

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:41:00 ; Search time 32.4667 Seconds
(without alignments)
53.778 Million cell updates/sec

Title: US-09-909-164-10

Perfect score: 52

Sequence: 1 BEVVFVXGMSYS(11)

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_19Jun03.*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|----------------------------|
| 1 | 50 | 96.2 | 11 | 23 | ABB80521 Hepatitis C virus |
| 2 | 50 | 96.2 | 11 | 23 | ABB80522 Hepatitis C virus |
| 3 | 50 | 96.2 | 11 | 23 | ABB80525 Hepatitis C virus |
| 4 | 50 | 96.2 | 11 | 23 | ABB80526 Hepatitis C virus |
| 5 | 50 | 96.2 | 11 | 23 | ABB80559 Hepatitis C virus |
| 6 | 50 | 96.2 | 11 | 23 | ABB80563 Hepatitis C virus |
| 7 | 50 | 96.2 | 11 | 23 | ABB80564 Hepatitis C virus |
| 8 | 50 | 96.2 | 11 | 23 | ABB80565 Hepatitis C virus |
| 9 | 50 | 96.2 | 11 | 23 | ABB80566 Hepatitis C virus |

| | | | | | |
|----|----|------|------|----|----------|
| 10 | 50 | 96.2 | 11 | 23 | ABB80567 |
| 11 | 50 | 96.2 | 11 | 23 | ABB80568 |
| 12 | 46 | 88.5 | 11 | 23 | ABB80524 |
| 13 | 46 | 88.5 | 11 | 23 | ABB80528 |
| 14 | 46 | 88.5 | 11 | 23 | ABB80529 |
| 15 | 46 | 88.5 | 11 | 23 | ABB80561 |
| 16 | 46 | 88.5 | 11 | 23 | ABB80562 |
| 17 | 45 | 86.5 | 11 | 23 | ABB80523 |
| 18 | 45 | 86.5 | 11 | 23 | ABB80527 |
| 19 | 45 | 86.5 | 11 | 23 | ABB80535 |
| 20 | 45 | 86.5 | 11 | 23 | ABB80536 |
| 21 | 45 | 86.5 | 11 | 23 | ABB80539 |
| 22 | 45 | 86.5 | 11 | 23 | ABB80540 |
| 23 | 45 | 86.5 | 11 | 23 | ABB80558 |
| 24 | 45 | 86.5 | 11 | 23 | ABB80560 |
| 25 | 44 | 84.6 | 11 | 23 | ABB80544 |
| 26 | 44 | 84.6 | 11 | 23 | ABB80545 |
| 27 | 44 | 84.6 | 11 | 23 | ABB80549 |
| 28 | 44 | 84.6 | 11 | 23 | ABB80552 |
| 29 | 44 | 84.6 | 11 | 23 | ABB80553 |
| 30 | 42 | 80.8 | 11 | 23 | ABB80530 |
| 31 | 41 | 78.8 | 11 | 23 | ABB80538 |
| 32 | 41 | 78.8 | 11 | 23 | ABB80542 |
| 33 | 41 | 78.8 | 11 | 23 | ABB80543 |
| 34 | 40 | 76.9 | 11 | 23 | ABB80537 |
| 35 | 40 | 76.9 | 11 | 23 | ABB80541 |
| 36 | 40 | 76.9 | 11 | 23 | ABB80547 |
| 37 | 40 | 76.9 | 11 | 23 | ABB80548 |
| 38 | 40 | 76.9 | 11 | 23 | ABB80551 |
| 39 | 40 | 76.9 | 11 | 23 | ABB80556 |
| 40 | 40 | 76.9 | 11 | 23 | ABB80557 |
| 41 | 40 | 76.9 | 20 | 20 | AAU76810 |
| 42 | 40 | 76.9 | 1022 | 22 | ABG03621 |
| 43 | 40 | 76.9 | 1022 | 22 | ABG05826 |
| 44 | 40 | 76.9 | 1022 | 22 | ABG08173 |
| 45 | 39 | 75.0 | 11 | 23 | ABB80546 |

ALIGNMENTS

RESULT 1

ABB80521

ID ABB80521 standard; peptide; 11 AA.

AC ABB80521;

XX

XX

08-OCT-2002 (first entry)

XX

Hepatitis C virus NS3/NS4a serine protease

XX

Hepatitis C virus; HCV; serine protease; i

XX

virucide.

XX

Synthetic.

XX

OS

XX

PH

Key

FT Modified-site

1

Location/Qualifiers

FT

Modified-site

6

/note= "N-terminal acetyl"

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

ALIGNMENTS

RESULT 1

ABB80521

ID ABB80521 standard; peptide; 11 AA.

AC ABB80521;

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #1.

KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;

KW virucide.

OS Synthetic.

PH Key Location/Qualifiers

FT Modified-site 1

FT Modified-site 6 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"

FT Modified-site 11 /note= "C-terminal amide"

FT WO200208251-A2.

PN 31-JAN-2002.

PD 19-JUL-2001; 2001WO-US23169.

PF 21-JUL-2000; 2000US-220101P.

PR

XX

```

PA (CORV-) CORVAS INT INC.
PI Lim-wilby M, Levy OE, Brunck TK;
XX
XX
XX
DR WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
XX Claim 17; Page 64; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
XX Sequence 11 AA;
SQ
Query Match 96.2%; Score 50; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 11
DB 1 EEVVPXGMSYS 11

RESULT 2
ABB80522
ID ABB80522 standard; peptide; 11 AA.
XX
XX ABB80522;
XX
XX 08-OCT-2002 (first entry)
XX
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #2.
XX
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX Synthetic.
XX
XX Key Location/Qualifiers
FH Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX
XX WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US23169.
XX
XX 21-JUL-2000; 2000US-220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
XX Claim 17; Page 64; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
XX Sequence 11 AA;
SQ
Query Match 96.2%; Score 50; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 11
DB 1 EEVVPXGMSYS 11

RESULT 3
ABB80525
ID ABB80525 standard; peptide; 11 AA.
XX
XX ABB80525;
XX
XX 08-OCT-2002 (first entry)
XX
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #5.
XX
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX Synthetic.
XX
XX Key Location/Qualifiers
FH Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX
XX WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US23169.
XX
XX 21-JUL-2000; 2000US-220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
XX Claim 17; Page 64; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of

```

CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.

XX Sequence 11 AA;

Query Match 96.2%; Score 50; DB 23; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.0013;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EEVVPXGMSYS 11
 DB 1 EEVVPXGMSYS 11
 |||||

RESULT 4
 ABB80526
 ID ABB80526 standard; peptide; 11 AA.
 XX
 AC ABB80526;
 DT 08-OCT-2002 (first entry)
 XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #6.

XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.
 XX Synthetic.

XX Key Location/Qualifiers
 FH Modified-site 1
 FT /note= "N-terminal acetyl"
 FT Modified-site 6
 FT /note= "Norvalyl carbonyl forming keto-amide linkage with
 residue 7"
 FT Misc-difference 8
 FT /note= "D-form residue"
 FT Misc-difference 9
 FT /note= "D-form residue"
 FT Modified-site 11
 FT /note= "C-terminal amide"

WO200208251-A2.
 31-JAN-2002.

19-JUL-2001; 2001WO-US23169.
 21-JUL-2000; 2000US-220101P.

(CORV-) CORVAS INT INC.

Lim-wilby M, Levy OE, Brunck TK;

WPI; 2002-361643/39.

Novel peptide compound having hepatitis C virus protease inhibitory
 activity useful for treating disorders associated with hepatitis C
 virus protease -

Claim 17; Page 64; 69pp; English.

The sequence represents a peptide compound of the invention having
 hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 invention are alpha-ketoamide peptide analogues. The peptides have
 virucide activity, and are useful for treating and in the manufacture of
 a medicament to treat disorders associated with HCV protease. A
 pharmaceutical composition comprising the peptide as an active ingredient
 is useful for treating disorders associated with hepatitis C virus.

XX Sequence 11 AA;

Query Match 96.2%; Score 50; DB 23; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.0013;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EEVVPXGMSYS 11
 DB 1 EEVVPXGMSYS 11
 |||||

RESULT 5
 ABB80559
 ID ABB80559 standard; peptide; 11 AA.
 XX
 AC ABB80559;

DT 08-OCT-2002 (first entry)

XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #39.
 XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.

XX Synthetic.

XX Key Location/Qualifiers
 FH Modified-site 1
 FT /note= "N-terminal acetyl"
 FT Modified-site 6
 FT /note= "Norvalyl carbonyl forming keto-amide linkage with
 residue 7"
 FT Misc-difference 8
 FT /note= "D-form residue"
 FT Modified-site 8
 FT /note= "Oxymethionine"
 FT Modified-site 11
 FT /note= "C-terminal amide"

WO200208251-A2.

31-JAN-2002.

19-JUL-2001; 2001WO-US23169.

21-JUL-2000; 2000US-220101P.

(CORV-) CORVAS INT INC.

Lim-wilby M, Levy OE, Brunck TK;

WPI; 2002-361643/39.

Novel peptide compound having hepatitis C virus protease inhibitory
 activity useful for treating disorders associated with hepatitis C
 virus protease -

Claim 17; Page 65; 69pp; English.

The sequence represents a peptide compound of the invention having
 hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 invention are alpha-ketoamide peptide analogues. The peptides have
 virucide activity, and are useful for treating and in the manufacture of
 a medicament to treat disorders associated with HCV protease. A
 pharmaceutical composition comprising the peptide as an active ingredient
 is useful for treating disorders associated with hepatitis C virus.

XX Sequence 11 AA;

Query Match 96.2%; Score 50; DB 23; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.0013;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EEVVPXGMSYS 11
 |||||

```

Db      1  EEVVPXGMSYS 11

RESULT 6
ABB80563
ID  ABB80563 standard; peptide; 11 AA.
XX
XX
AC  ABB80563;
XX
DT  08-OCT-2002 (first entry)
XX
DE  Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #43.
XX
KW  Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW  virucide.
XX
OS  Synthetic.
XX
FH  Key
FT  Modified-site 1 Location/Qualifiers
FT  Modified-site 6 /note= "N-terminal acetyl"
FT  Modified-site 6 /note= "Leucyl carbonyl forming keto-amide linkage with
FT  Modified-site 11 residue 7"
FT  Modified-site 11 /note= "C-terminal amide"
XX
PN  WO200208251-A2.
XX
XX  31-JAN-2002.
XX
XX  19-JUL-2001; 2001WO-US23169.
XX
XX  21-JUL-2000; 2000US-220101P.
XX
XX  (CORV-) CORVAS INT INC.
XX
XX  Lim-wilby M, Levy OE, Brunck TK;
XX
XX  WPI; 2002-361643/39.
XX
XX  Novel peptide compound having hepatitis C virus protease inhibitory
PT  activity useful for treating disorders associated with hepatitis C
PT  virus protease -
XX
XX  Claim 17; Page 65; 69pp; English.
XX
XX  The sequence represents a peptide compound of the invention having
CC  hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC  invention are alpha-ketoamide peptide analogues. The peptides have
CC  virucide activity, and are useful for treating and in the manufacture of
CC  a medicament to treat disorders associated with HCV protease. A
CC  pharmaceutical composition comprising the peptide as an active ingredient
CC  is useful for treating disorders associated with hepatitis C virus.
XX
XX  Sequence 11 AA;
XX
XX  Query Match 96.2%; Score 50; DB 23; Length 11;
XX  Best Local Similarity 100.0%; Pred. No. 0.0013;
XX  Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY  1  EEVVPXGMSYS 11
    |||||
DB  1  EEVVPXGMSYS 11

RESULT 8
ABB80565
ID  ABB80565 standard; peptide; 11 AA.
XX
XX  ABB80565;
XX
DT  08-OCT-2002 (first entry)
XX
DE  Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #45.
XX
KW  Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW  virucide.
XX
OS  Synthetic.
XX
XX  Key
FT  Modified-site 1 Location/Qualifiers
FT  Modified-site 11 /note= "N-terminal acetyl"
FT  Modified-site 11 /note= "Leucyl carbonyl forming keto-amide linkage with
FT  Modified-site 11 residue 7"
FT  Modified-site 11 /note= "C-terminal amide"
XX
PN  WO200208251-A2.
XX
XX  31-JAN-2002.
XX
XX  19-JUL-2001; 2001WO-US23169.
XX
XX  21-JUL-2000; 2000US-220101P.
XX
XX  (CORV-) CORVAS INT INC.
XX
XX  Lim-wilby M, Levy OE, Brunck TK;
XX
XX  WPI; 2002-361643/39.
XX
XX  Novel peptide compound having hepatitis C virus protease inhibitory
PT  activity useful for treating disorders associated with hepatitis C
PT  virus protease -
XX
XX  Claim 17; Page 65; 69pp; English.
XX
XX  The sequence represents a peptide compound of the invention having
CC  hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC  invention are alpha-ketoamide peptide analogues. The peptides have
CC  virucide activity, and are useful for treating and in the manufacture of
CC  a medicament to treat disorders associated with HCV protease. A
CC  pharmaceutical composition comprising the peptide as an active ingredient
CC  is useful for treating disorders associated with hepatitis C virus.
XX
XX  Sequence 11 AA;
XX
XX  Query Match 96.2%; Score 50; DB 23; Length 11;
XX  Best Local Similarity 100.0%; Pred. No. 0.0013;
XX  Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY  1  EEVVPXGMSYS 11
    |||||
DB  1  EEVVPXGMSYS 11

RESULT 7
ABB80564
ID  ABB80564 standard; peptide; 11 AA.
XX
XX  ABB80564;
XX
DT  08-OCT-2002 (first entry)
XX

```



```

FT FT /note= "N-terminal acetyl"
FT FT 6
FT FT /note= "Norleucyl carbonyl forming keto-amide linkage
FT FT with residue 7"
FT FT 11
FT FT /note= "C-terminal amide"
XX XX
PN WO200208251-A2.
XX XX
XX 31-JAN-2002.
XX XX
PF 19-JUL-2001; 2001WO-US23169.
XX XX
XX 21-JUL-2000; 2000US-220101P.
XX XX
XX (CORV-) CORVAS INT INC.
XX XX
XX Lim-wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX XX
XX Claim 17; Page 65; 69pp; English.
XX XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX XX
SQ Sequence 11 AA;

Query Match 96.2%; Score 50; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11
DB 1 BEVVPXGMSYS 11

RESULT 9
ABB80566
ID ABB80566 standard; peptide; 11 AA.
XX AC ABB80566;
XX XX
XX 08-OCT-2002 (first entry)
XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #46.
XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX OS Synthetic.
XX FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "2-aminoisobutyl carbonyl residue forming a
FT keto-amide linkage with residue 7"
FT Modified-site 11 /note= "C-terminal amide"
XX XX
XX WO200208251-A2.
XX 31-JAN-2002.
XX XX
XX 19-JUL-2001; 2001WO-US23169.
XX PF
XX 21-JUL-2000; 2000US-220101P.
XX PR
XX (CORV-) CORVAS INT INC.
XX PA
XX Lim-wilby M, Levy OE, Brunck TK;
XX PI
XX WPI; 2002-361643/39.
XX DR

```

```

XX 19-JUL-2001; 2001WO-US23169.
XX PF
XX 21-JUL-2000; 2000US-220101P.
XX PR
XX (CORV-) CORVAS INT INC.
XX PA
XX Lim-wilby M, Levy OE, Brunck TK;
XX PI
XX WPI; 2002-361643/39.
XX DR
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX XX
XX Claim 17; Page 65; 69pp; English.
XX PS
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX XX
SQ Sequence 11 AA;

Query Match 96.2%; Score 50; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11
DB 1 BEVVPXGMSYS 11

RESULT 10
ABB80567
ID ABB80567 standard; peptide; 11 AA.
XX AC ABB80567;
XX XX
XX 08-OCT-2002 (first entry)
XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #47.
XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX OS Synthetic.
XX FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "(s,s)allothreonyl carbonyl residue forming a
FT keto-amide linkage with residue 7"
FT Modified-site 11 /note= "C-terminal amide"
XX XX
XX WO200208251-A2.
XX 31-JAN-2002.
XX XX
XX 19-JUL-2001; 2001WO-US23169.
XX PF
XX 21-JUL-2000; 2000US-220101P.
XX PR
XX (CORV-) CORVAS INT INC.
XX PA
XX Lim-wilby M, Levy OE, Brunck TK;
XX PI
XX WPI; 2002-361643/39.
XX DR

```

XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
PS Claim 17; Page 65; 69pp; English.
XX
CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;
Query Match 96.2%; Score 50; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EEVVPXGMSYS 11
DB 1 EEVVPXGMSYS 11
RESULT 11
ABB80568
ID ABB80568 standard; peptide; 11 AA.
XX
AC ABB80568;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #48.
XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Alpha-propynyl-glycyl-carbonyl residue forming
FT a keto-amide linkage with residue 7"
FT Modified-site 11 /note= "C-terminal amide"
FT WO200208251-A2.
XX
PD 31-JAN-2002.
XX
PF 19-JUL-2001; 2001WO-US23169.
XX
PR 21-JUL-2000; 2000US-220101P.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Lim-wilby M, Levy OE, Brunck TK;
XX
DR WPI; 2002-361643/39.
XX
PT Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
XX Claim 17; Page 65; 69pp; English.
XX
CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
XX virucide activity, and are useful for treating and in the manufacture of
XX a medicament to treat disorders associated with HCV protease. A
XX pharmaceutical composition comprising the peptide as an active ingredient
XX is useful for treating disorders associated with hepatitis C virus.

CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;
Query Match 96.2%; Score 50; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EEVVPXGMSYS 11
DB 1 EEVVPXGMSYS 11
RESULT 12
ABB80524
ID ABB80524 standard; peptide; 11 AA.
XX
AC ABB80524;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #4.
XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
FT WO200208251-A2.
XX
PD 31-JAN-2002.
XX
PF 19-JUL-2001; 2001WO-US23169.
XX
PR 21-JUL-2000; 2000US-220101P.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Lim-wilby M, Levy OE, Brunck TK;
XX
DR WPI; 2002-361643/39.
XX
PT Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
XX Claim 17; Page 64; 69pp; English.
XX
CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;
Query Match 88.5%; Score 46; DB 23; Length 11;

Best Local Similarity 90.9%; Pred. No. 0.0088; Mismatches 1; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGMSYS 11
Db |||||
1 EEVVPXGMDYS 11

RESULT 13
ABB80528
ID ABB80528 standard; peptide; 11 AA.
XX ABB80528;
XX
XX 08-OCT-2002 (first entry)
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #8.
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX Synthetic.
XX
XX Key Location/Qualifiers
FH Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX WO200208251-A2.
XX
XX 31-JAN-2002.
XX 19-JUL-2001; 2001WO-US23169.
XX 21-JUL-2000; 2000US-220101P.
XX (CORV-) CORVAS INT INC.
XX Lim-wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX Claim 17; Page 64; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX Sequence 11 AA;

Query Match 88.5%; Score 46; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0088;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGMSYS 11
Db |||||
1 EEVVPXGMDYS 11

RESULT 15
ABB80561
ID ABB80561 standard; peptide; 11 AA.
XX
XX ABB80561;
XX

RESULT 14
ABB80529
ID ABB80529 standard; peptide; 11 AA.
XX ABB80529;
XX
XX 08-OCT-2002 (first entry)
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #9.
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX Synthetic.
XX
XX Key Location/Qualifiers
FH Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX WO200208251-A2.
XX
XX 31-JAN-2002.
XX 19-JUL-2001; 2001WO-US23169.
XX 21-JUL-2000; 2000US-220101P.
XX (CORV-) CORVAS INT INC.
XX Lim-wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX Claim 17; Page 64; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX Sequence 11 AA;

Query Match 88.5%; Score 46; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0088;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGMSYS 11
Db |||||
1 EEVVPXGMDYS 11

RESULT 15
ABB80561
ID ABB80561 standard; peptide; 11 AA.
XX
XX ABB80561;
XX

DT 08-OCT-2002 (first entry)
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #41.
DE
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX
XX Synthetic.
OS
XX
FF Key Location/Qualifiers
FT Modified-site 1
FT /note= "N-terminal acetyl"
FT Modified-site 6
FT /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 8
FT /note= "D-form residue"
FT Modified-site 8
FT /note= "Oxymethionine"
FT Modified-site 11
FT /note= "C-terminal amide"
XX
PN WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US23169.
PF
XX 21-JUL-2000; 2000US-220101P.
PR
XX (CORV-) CORVAS INT INC.
PA
XX Lim-wilby M, Levy OE, Brunck TK;
PI
XX WPI; 2002-361643/39.
DR
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
XX Claim 17; Page 65; 69pp; English.
PS
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;
Query Match 88.5%; Score 46; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0088;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 EEVVPXGMSYS 11
| | | | | | | |
Db 1 EEVVPXGMDYS 11

Search completed: December 22, 2003, 17:41:01
Job time : 33.4667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:37:03 ; Search time 10.1333 Seconds
(without alignments)
45.930 Million cell updates/sec

Title: US-09-909-164-10
Perfect score: 52
Sequence: 1 EVVPMXGMSYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/ECTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 34 | 65.4 | 947 | 4 | US-09-228-986-73 |
| 2 | 34 | 65.4 | 1407 | 4 | Sequence 73, Appl |
| 3 | 33 | 63.5 | 45 | 2 | Sequence 7885, Ap |
| 4 | 33 | 63.5 | 45 | 2 | Sequence 236, App |
| 5 | 33 | 63.5 | 45 | 3 | Sequence 236, App |
| 6 | 33 | 63.5 | 45 | 4 | Sequence 236, App |
| 7 | 33 | 63.5 | 65 | 6 | Sequence 236, App |
| 8 | 33 | 63.5 | 410 | 6 | Patent No. 5177197 |
| 9 | 32 | 61.5 | 1394 | 6 | Patent No. 5177197 |
| 10 | 32 | 61.5 | 10 | 4 | Sequence 66, Appl |
| 11 | 32 | 61.5 | 10 | 4 | Sequence 66, Appl |
| 12 | 32 | 61.5 | 102 | 2 | Sequence 66, Appl |
| 13 | 32 | 61.5 | 152 | 2 | Sequence 23, Appl |
| 14 | 32 | 61.5 | 152 | 3 | Sequence 4, Appl |
| 15 | 32 | 61.5 | 152 | 3 | Sequence 4, Appl |
| 16 | 32 | 61.5 | 173 | 1 | Sequence 4, Appl |
| 17 | 32 | 61.5 | 189 | 2 | Sequence 7, Appl |
| 18 | 32 | 61.5 | 189 | 2 | Sequence 21, Appl |
| 19 | 32 | 61.5 | 189 | 3 | Sequence 21, Appl |
| 20 | 32 | 61.5 | 189 | 5 | Sequence 21, Appl |
| 21 | 32 | 61.5 | 236 | 2 | Sequence 21, Appl |
| 22 | 32 | 61.5 | 236 | 2 | Sequence 22, Appl |
| 23 | 32 | 61.5 | 236 | 3 | Sequence 22, Appl |
| 24 | 32 | 61.5 | 236 | 5 | Sequence 22, Appl |
| 25 | 32 | 61.5 | 280 | 2 | Sequence 6, Appl |
| 26 | 32 | 61.5 | 280 | 3 | Sequence 6, Appl |
| 27 | 32 | 61.5 | 289 | 2 | Sequence 4, Appl |

28 32 61.5 289 5 PCT-US93-05000-4 Sequence 4, Appli
29 32 61.5 291 5 PCT-US93-05000-6 Sequence 6, Appli
30 32 61.5 292 2 US-08-464-517-23 Sequence 23, Appl
31 32 61.5 292 2 US-08-246-361A-6 Sequence 6, Appli
32 32 61.5 292 2 US-08-246-361A-23 Sequence 23, Appl
33 32 61.5 292 3 US-08-463-772-23 Sequence 23, Appl
34 32 61.5 292 5 PCT-US93-05000-23 Sequence 23, Appl
35 32 61.5 295 1 US-07-947-120-8 Sequence 8, Appli
36 32 61.5 295 1 US-08-472-893A-8 Sequence 8, Appli
37 32 61.5 295 2 US-08-460-694-2 Sequence 2, Appli
38 32 61.5 295 2 US-08-464-517-19 Sequence 19, Appl
39 32 61.5 295 2 US-08-464-517-20 Sequence 20, Appl
40 32 61.5 295 2 US-08-246-361A-19 Sequence 19, Appl
41 32 61.5 295 2 US-08-246-361A-20 Sequence 20, Appl
42 32 61.5 295 3 US-08-463-772-19 Sequence 19, Appl
43 32 61.5 295 3 US-08-463-772-20 Sequence 20, Appl
44 32 61.5 295 3 US-08-460-744-2 Sequence 2, Appli
45 32 61.5 295 3 US-07-667-711B-2 Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-228-986-73
; Sequence 73, Application US/09228986
; Patent No. 6359198
; GENERAL INFORMATION:
; APPLICANT: Strabala, Timothy
; APPLICANT: Nieuwenhuizen, Niels
; TITLE OF INVENTION: Compositions Isolated from Plant Cells
; TITLE OF INVENTION: and Their Use in the Modification of Plant Cell Signalling
; FILE REFERENCE: 11000/1020
; CURRENT APPLICATION NUMBER: US/09/228,986
; CURRENT FILING DATE: 1999-01-12
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 73
; LENGTH: 947
; TYPE: PRT
; ORGANISM: Pinus radiata
US-09-228-986-73

Query Match 65.4%; Score 34; DB 4; Length 947;
Best Local Similarity 66.7%; Pred. No. 1.3e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 3 VVPXGMSYS 11
|:|:|:|:
Db 686 VMPGSGISYS 694

RESULT 2
US-09-328-352-7885
; Sequence 7885, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7885
; LENGTH: 1407
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7885

Query Match 65.4%; Score 34; DB 4; Length 1407;
Best Local Similarity 66.7%; Pred. No. 2.1e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

Qy      2 BEVPXGMSY 10
Db      596 BEVPEGLSF 604

RESULT 3
US-08-637-759B-236
; Sequence 236, Application US/08637759B
; Patent No. 5876931
; GENERAL INFORMATION:
; APPLICANT: David William Holden
; TITLE OF INVENTION: Identification of Genes
; NUMBER OF SEQUENCES: 501
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/637.759B
; FILING DATE: 03-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB95/02875
; FILING DATE: 11-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: RPMS 101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 236:
; LENGTH: 45 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; US-08-637-759B-236

Query Match      63.5%; Score 33; DB 2; Length 45;
Best Local Similarity 60.0%; Pred. No. 6.6;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy      1 BEVPXGMSY 10
Db      1 BEISPLGWSY 10

RESULT 4
US-08-871-355A-236
; Sequence 236, Application US/08871355A
; Patent No. 6015669
; GENERAL INFORMATION:
; APPLICANT: David William Holden
; TITLE OF INVENTION: Identification of Genes
; NUMBER OF SEQUENCES: 501
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/201.945
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/637,759
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:

```

```

; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/871.355A
; FILING DATE: 09-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB95/02875
; FILING DATE: 11-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: RPMS 101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 236:
; LENGTH: 45 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; US-08-871-355A-236

Query Match      63.5%; Score 33; DB 3; Length 45;
Best Local Similarity 60.0%; Pred. No. 6.6;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy      1 BEVPXGMSY 10
Db      1 BEISPLGWSY 10

RESULT 5
US-09-201-945-236
; Sequence 236, Application US/09201945
; Patent No. 6342215
; GENERAL INFORMATION:
; APPLICANT: David William Holden
; TITLE OF INVENTION: Identification of Genes
; NUMBER OF SEQUENCES: 501
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/201.945
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/637,759
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:

```

; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: RPMS 101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 236:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 45 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
US-09-201-945-236

Query Match 63.5%; Score 33; DB 4; Length 45;
Best Local Similarity 60.0%; Pred. No. 6.6;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMSY 10
||: ||: ||:
DB 1 BEISPLGWSY 10

RESULT 6

; Patent No. 5177197
; APPLICANT: KANZAKI, TETSUTO;OLOFSSON, ANDERS;MOREN, ANITA;
; WERNSTEDT, CHRISTER;HELLMAN, ULF;MIYAZONO, KOHEI;CLAESSON-WELSH,
; LENA;HELDIN, CARL-HENRIK
; TITLE OF INVENTION: ISOLATED NUCLEOTIDE SEQUENCE EXPRESSING
; HUMAN TRANSFORMING GROWTH FACTOR-BETAL-BINDING PROTEIN
; NUMBER OF SEQUENCES: 53
; CURRENT APPLICATION DATA:
; FILING DATE: 27-FEB-1990
; SEQ ID NO:51:
; LENGTH: 65
5177197-51

Query Match 63.5%; Score 33; DB 6; Length 65;
Best Local Similarity 45.5%; Pred. No. 10;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11
||: ||: ||:
DB 52 KEICPGMGY 62

RESULT 7

; Patent No. 5177197
; APPLICANT: KANZAKI, TETSUTO;OLOFSSON, ANDERS;MOREN, ANITA;
; WERNSTEDT, CHRISTER;HELLMAN, ULF;MIYAZONO, KOHEI;CLAESSON-WELSH,
; LENA;HELDIN, CARL-HENRIK
; TITLE OF INVENTION: ISOLATED NUCLEOTIDE SEQUENCE EXPRESSING
; HUMAN TRANSFORMING GROWTH FACTOR-BETAL-BINDING PROTEIN
; NUMBER OF SEQUENCES: 53
; CURRENT APPLICATION DATA:
; FILING DATE: 27-FEB-1990
; SEQ ID NO:1:
; LENGTH: 410
5177197-1

Query Match 63.5%; Score 33; DB 6; Length 410;
Best Local Similarity 45.5%; Pred. No. 81;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11
||: ||: ||:
DB 399 KEICPGMGY 409

RESULT 8

; Patent No. 5177197
; APPLICANT: KANZAKI, TETSUTO;OLOFSSON, ANDERS;MOREN, ANITA;
; WERNSTEDT, CHRISTER;HELLMAN, ULF;MIYAZONO, KOHEI;CLAESSON-WELSH,
; LENA;HELDIN, CARL-HENRIK
; TITLE OF INVENTION: ISOLATED NUCLEOTIDE SEQUENCE EXPRESSING
; HUMAN TRANSFORMING GROWTH FACTOR-BETAL-BINDING PROTEIN
; NUMBER OF SEQUENCES: 53
; CURRENT APPLICATION DATA:
; FILING DATE: 27-FEB-1990
; SEQ ID NO:30:
; LENGTH: 1394
5177197-30

Query Match 63.5%; Score 33; DB 6; Length 1394;
Best Local Similarity 45.5%; Pred. No. 3.3e+02;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11
||: ||: ||:
DB 399 KEICPGMGY 409

RESULT 9

; US-09-357-952-66
; Sequence 66, Application US/09357952
; Patent No. 6248904
; GENERAL INFORMATION:
; APPLICANT: Zhang, Han-Zhong
; APPLICANT: Cai, Sui Xiong
; APPLICANT: Drewe, John A.
; APPLICANT: Yang, Wu
; TITLE OF INVENTION: Fluorescence Screening Assays for Caspases, Proteases
; TITLE OF INVENTION: Other Enzymes and the Use Thereof
; FILE REFERENCE: 1735.0030001
; CURRENT APPLICATION NUMBER: US/09/357,952
; CURRENT FILING DATE: 1999-07-21
; EARLIER APPLICATION NUMBER: US 60/093,642
; EARLIER FILING DATE: 21-JUL-1998
; NUMBER OF SEQ ID NOS: 139
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 66
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Synthetic
; OTHER INFORMATION: Peptide
US-09-357-952-66

Query Match 61.5%; Score 32; DB 3; Length 10;
Best Local Similarity 50.0%; Pred. No. 1.9;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMSY 10
||: ||: ||:
DB 1 DDIVPCMSY 10

RESULT 10

; US-09-521-650-66
; Sequence 66, Application US/09521650
; Patent No. 6335429
; GENERAL INFORMATION:
; APPLICANT: Weber, Eckard
; APPLICANT: Cai, Sui Xiong
; APPLICANT: Keana, John F.W.
; APPLICANT: Drewe, John A.

APPLICANT: Zhang, Han-Zhong
TITLE OF INVENTION: No. 6335429el Fluorogenic or Fluorescent Reporter Molecules and
TITLE OF INVENTION: Their Applications for Whole-Cell Fluorescence
TITLE OF INVENTION: Screening Assays for Caspases and Other Enzymes and the
TITLE OF INVENTION: Use Thereof
FILE REFERENCE: 1735.0290002
CURRENT APPLICATION NUMBER: US/09/521,650
CURRENT FILING DATE: 2000-03-08
EARLIER APPLICATION NUMBER: 09/168,888
EARLIER FILING DATE: 1998-10-09
EARLIER APPLICATION NUMBER: US 60/061,582
EARLIER FILING DATE: 1997-10-10
EARLIER APPLICATION NUMBER: US 09/033,661
EARLIER FILING DATE: 1998-03-03
NUMBER OF SEQ ID NOS: 142
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 66
LENGTH: 10
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic
OTHER INFORMATION: Peptide
US-09-521-650-66

Query Match 61.5%; Score 32; DB 4; Length 10;
Best Local Similarity 50.0%; Pred. No. 1.9;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMSY 10
:::|::|::|
Db 1 DDIVPCSMYS 10

RESULT 11

US-09-168-888-66
Sequence 66, Application US/09168888
Patent No. 6342611

GENERAL INFORMATION:
APPLICANT: Weber, Eckard

APPLICANT: Cai, Sui Xiong

APPLICANT: Keana, John F.W.

APPLICANT: Drewe, John A.

APPLICANT: Zhang, Han-Zhong

TITLE OF INVENTION: No. 6342611el Fluorogenic or Fluorescent Reporter Molecules and
TITLE OF INVENTION: Their Applications for Whole-Cell Fluorescence

TITLE OF INVENTION: Screening Assays for Caspases and Other Enzymes and the

TITLE OF INVENTION: Use Thereof

FILE REFERENCE: 1735.0290002

CURRENT APPLICATION NUMBER: US/09/168,888

CURRENT FILING DATE: 1998-10-09

EARLIER APPLICATION NUMBER: US 60/061,582

EARLIER FILING DATE: 1997-10-10

EARLIER APPLICATION NUMBER: US 09/033,661

EARLIER FILING DATE: 1998-03-03

NUMBER OF SEQ ID NOS: 142

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 66

LENGTH: 10

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: Synthetic

OTHER INFORMATION: Peptide

US-09-168-888-66

Query Match 61.5%; Score 32; DB 4; Length 10;
Best Local Similarity 50.0%; Pred. No. 1.9;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMSY 10
:::|::|::|
Db 1 DDIVPCSMYS 10

RESULT 12

US-08-580-988A-23

Sequence 23, Application US/08580988A

Patent No. 5856161

GENERAL INFORMATION:

APPLICANT: Aggarwal et al.

TITLE OF INVENTION: Tumor Necrosis Factor

TITLE OF INVENTION: Receptor-I-Associated Protein Kinase And Methods

TITLE OF INVENTION: For Its Use

NUMBER OF SEQUENCES: 27

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dr. Benjamin A. Adler

STREET: 8011 Candle Lane

CITY: Houston

STATE: Texas

COUNTRY: USA

ZIP: 77071

COMPUTER READABLE FORM:

MEDIUM TYPE: 1.44 Mb floppy disk

COMPUTER: Apple Macintosh

OPERATING SYSTEM: Macintosh

SOFTWARE: Microsoft Word for Macintosh

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/580,988A

FILING DATE: January 3, 1996

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Benjamin Aaron Adler, Ph.D., J.D.

REGISTRATION NUMBER: 35,423

REFERENCE/DOCKET NUMBER: D5721CIP2

TELECOMMUNICATION INFORMATION:

TELEPHONE: 713-777-2321

TELEFAX: 713-777-6908

INFORMATION FOR SEQ ID NO: 23:

SEQUENCE CHARACTERISTICS:

LENGTH: 102 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE:

DESCRIPTION: protein

HYPOTHETICAL: no

ANTI-SENSE: no

FRAGMENT TYPE: internal

ORIGINAL SOURCE:

US-08-580-988A-23

Query Match 61.5%; Score 32; DB 2; Length 102;
Best Local Similarity 60.0%; Pred. No. 27;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMSY 10
|::|::|::|

Db 24 BEVPLAWNY 33

RESULT 13

US-08-460-694-4

Sequence 4, Application US/08460694

Patent No. 5858655

GENERAL INFORMATION:

APPLICANT: Arnold, Andrew

TITLE OF INVENTION: PRAD1 Cyclin and its cDNA

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.

STREET: 1100 New York Avenue, N.W., Suite 600

CITY: Washington

Query Match 61.5%; Score 32; DB 4; Length 10;
Best Local Similarity 50.0%; Pred. No. 1.9;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMSY 10
:::|::|::|
Db 1 DDIVPCSMYS 10

STATS: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,694
FILING DATE: 02-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McConathy, Evelyn H.
REGISTRATION NUMBER: 35,279
REFERENCE/DOCKET NUMBER: 0609.4070002
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 152 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-460-694-4

Query Match 61.5%; Score 32; DB 2; Length 152;
Best Local Similarity 60.0%; Pred. No. 42;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVVPXGMSY 10
||| | | |
Db 20 BEVFPPLAMNY 29

RESULT 14
US-08-460-744-4
Sequence 4, Application US/08460744
Patent No. 6107541
GENERAL INFORMATION:
APPLICANT: Arnold, Andrew
TITLE OF INVENTION: Prad1 Cyclin and its cDNA
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 New York Avenue, N.W., Suite 600
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,744
FILING DATE: 02-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McConathy, Evelyn H.
REGISTRATION NUMBER: 35,279
REFERENCE/DOCKET NUMBER: 0609.4070005
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 152 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant

TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-460-744-4

Query Match 61.5%; Score 32; DB 3; Length 152;
Best Local Similarity 60.0%; Pred. No. 42;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVVPXGMSY 10
||| | | |
Db 20 BEVFPPLAMNY 29

RESULT 15
US-07-667-711B-4
Sequence 4, Application US/07667711B
Patent No. 6110700
GENERAL INFORMATION:
APPLICANT: ARNOLD, ANDREW
TITLE OF INVENTION: Prad1 Cyclin and its cDNA
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVE., NW, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/667,711B
FILING DATE: 11-MAR-1991
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MCPHAIL, DONALD R.
REGISTRATION NUMBER: 35,811
REFERENCE/DOCKET NUMBER: 0609.4070000
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 152 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
US-07-667-711B-4

Query Match 61.5%; Score 32; DB 3; Length 152;
Best Local Similarity 60.0%; Pred. No. 42;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVVPXGMSY 10
||| | | |
Db 20 BEVFPPLAMNY 29

Search completed: December 22, 2003, 16:43:43
Job time : 10.1333 secs

| Result No. | Score | Query | | Length | DB | ID | Description |
|------------|-------|-------|------|--------|---------------------|----|--------------------|
| | | Match | | | | | |
| 1 | 38 | 73.1 | 3472 | 12 | US-10-029-120-4 | | Sequence 4, Appli |
| 2 | 38 | 73.1 | 3472 | 14 | US-10-027-806-4 | | Sequence 4, Appli |
| 3 | 38 | 73.1 | 3472 | 14 | US-10-034-623-4 | | Sequence 4, Appli |
| 4 | 38 | 73.1 | 3472 | 15 | US-10-027-801-4 | | Sequence 4, Appli |
| 5 | 36 | 69.2 | 153 | 12 | US-10-029-386-32486 | | Sequence 32486, A |
| 6 | 35 | 67.3 | 236 | 12 | US-10-029-386-32076 | | Sequence 32076, A |
| 7 | 34 | 65.4 | 290 | 12 | US-10-094-749-2565 | | Sequence 2565, Ap |
| 8 | 34 | 65.4 | 947 | 15 | US-10-101-4648-73 | | Sequence 73, Appli |
| 9 | 34 | 65.4 | 1163 | 12 | US-10-122-067-4 | | Sequence 4, Appli |
| 10 | 34 | 65.4 | 1499 | 12 | US-10-096-534-67 | | Sequence 67, Appli |
| 11 | 34 | 65.4 | 1499 | 12 | US-10-102-067-2 | | Sequence 2, Appli |
| 12 | 33 | 63.5 | 99 | 15 | US-10-106-698-7477 | | Sequence 7477, Ap |
| 13 | 33 | 63.5 | 426 | 15 | US-10-214-766-43 | | Sequence 43, Appli |
| 14 | 33 | 63.5 | 478 | 11 | US-09-992-600A-108 | | Sequence 108, App |
| 15 | 33 | 63.5 | 478 | 11 | US-09-924-340-108 | | Sequence 108, App |

```

; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCCRP.002A
; CURRENT APPLICATION NUMBER: US/10/027,806
; CURRENT FILING DATE: 2001-12-21
; PRIOR FILING DATE: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-027-806-4

Query Match      73.1%; Score 38; DB 14; Length 3472;
Best Local Similarity 54.5%; Pred. No. 1.6e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy      1 BEVVPXGMSYS 11
Db      2294 EDVIPRGISFS 2304

RESULT 3
US-10-034-623-4
; Sequence 4, Application US/10034623
; Publication No. US20020198365A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCCRP.002A
; CURRENT APPLICATION NUMBER: US/10/034,623
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: 60/102,294
; PRIOR FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-034-623-4

Query Match      73.1%; Score 38; DB 14; Length 3472;
Best Local Similarity 54.5%; Pred. No. 1.6e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy      1 BEVVPXGMSYS 11
Db      2294 EDVIPRGISFS 2304

RESULT 4
US-10-027-801-4
; Sequence 4, Application US/10027801
; Publication No. US20030054364A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCCRP.002A
; CURRENT APPLICATION NUMBER: US/10/027,801
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSEQ for Windows Version 3.0

```

```

; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-027-801-4

Query Match      73.1%; Score 38; DB 15; Length 3472;
Best Local Similarity 54.5%; Pred. No. 1.6e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy      1 BEVVPXGMSYS 11
Db      2294 EDVIPRGISFS 2304

RESULT 5
US-10-029-386-32486
; Sequence 32486, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 32486
; LENGTH: 153
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AF002994.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
; OTHER INFORMATION: SWISSPROT HIT: P43363, EVALUAE 5.00e-46
US-10-029-386-32486

Query Match      69.2%; Score 36; DB 12; Length 153;
Best Local Similarity 70.0%; Pred. No. 12;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy      1 BEVVPXGMSY 10
Db      15 KEVVPYCHSY 24

RESULT 6
US-10-029-386-32076
; Sequence 32076, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 32076
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:

```

```

; OTHER INFORMATION: MAP TO 298050.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.6
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.3
; OTHER INFORMATION: SWISSPROT HIT: P15822, EVALUATE 1.00e-125
US-10-029-386-32076

```

```

Query Match      67.3%; Score 35; DB 12; Length 236;
Best Local Similarity 66.7%; Pred. No. 32;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      3 VVPXGMSYS 11
Db      80 VVPAGLTYS 88

```

```

RESULT 7
US-10-094-749-2565
; Sequence 2565, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOKUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 2565
; LENGTH: 290
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-2565

```

```

Query Match      65.4%; Score 34; DB 12; Length 290;
Best Local Similarity 66.7%; Pred. No. 66;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      1 BEVVPXGMS 9
Db      35 BEIVPMGIS 43

```

```

RESULT 8
US-10-101-464A-73
; Sequence 73, Application US/10101464A
; Publication No. US20030046728A1
; GENERAL INFORMATION:

```

```

; APPLICANT: Strabala, Timothy
; APPLICANT: Nieuwenhuizen, Nicolaas
; APPLICANT: Higgins, Colleen M.
; TITLE OF INVENTION: Compositions Isolated from Plant Cells
; TITLE OF INVENTION: and Their Use in the Modification of Plant Cell Signaling
; FILE REFERENCE: 11000.1020c2
; CURRENT APPLICATION NUMBER: US/10/101,464A
; CURRENT FILING DATE: 2002-03-18
; PRIOR APPLICATION NUMBER: 09/704,302
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 09/228,986
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/162,866
; PRIOR FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PCT/US00/00724
; PRIOR FILING DATE: 2000-01-11
; NUMBER OF SEQ ID NOS: 989
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 73
; LENGTH: 947
; TYPE: PRT
; ORGANISM: Pinus radiata
US-10-101-464A-73

```

```

Query Match      65.4%; Score 34; DB 15; Length 947;
Best Local Similarity 66.7%; Pred. No. 2.5e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      3 VVPXGMSYS 11
Db      686 VMPGIGSYS 694

```

```

RESULT 9
US-10-122-067-4
; Sequence 4, Application US/10122067
; Publication No. US20030165883A1
; GENERAL INFORMATION:
; APPLICANT: ROY, A.J. Curtis
; TITLE OF INVENTION: 27091, A PHOSPHOLIPID TRANSPORTING
; TITLE OF INVENTION: ATPASE MOLECULE AND USES THEREFOR
; FILE REFERENCE: MP101-058PIR
; CURRENT APPLICATION NUMBER: US/10/122,067
; CURRENT FILING DATE: 2002-04-12
; PRIOR APPLICATION NUMBER: 60/283434
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0;
; SEQ ID NO 4
; LENGTH: 1163
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-122-067-4

```

```

Query Match      65.4%; Score 34; DB 12; Length 1163;
Best Local Similarity 72.7%; Pred. No. 3.1e+02;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

Qy      1 BEVVPXGMSYS 11
Db      133 BEVVPRGGSVS 143

```

```

RESULT 10
US-10-096-534-67
; Sequence 67, Application US/10096534
; Publication No. US20030166887A1
; GENERAL INFORMATION:
; APPLICANT: The Brigham and Women's Hospital, Inc.
; APPLICANT: Yates, Karen
; APPLICANT: Mizuno, Shuichi
; APPLICANT: Glowacki, Julie
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF SKELETAL DEGENERATION CONDITIONS

```

; FILE REFERENCE: B0801/7244/KA/ERP
 ; CURRENT APPLICATION NUMBER: US/10/096,534
 ; CURRENT FILING DATE: 2002-03-12
 ; PRIOR APPLICATION NUMBER: US 60/274,980
 ; PRIOR FILING DATE: 2001-03-12
 ; NUMBER OF SEQ ID NOS: 79
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 67
 ; LENGTH: 1499
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-096-534-67

Query Match 65.4%; Score 34; DB 12; Length 1499;
 Best Local Similarity 72.7%; Pred. No. 4.2e+02;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVVPKXGMSYS 11
 Db 469 BEVVRGGGSVS 479

RESULT 11

US-10-122-067-2
 ; Sequence 2, Application US/10122067
 ; Publication No. US20030165883A1
 ; GENERAL INFORMATION:

; APPLICANT: RORY A.J. CURTIS
 ; TITLE OF INVENTION: 27091, A PHOSPHOLIPID TRANSPORTING
 ; TITLE OF INVENTION: APTASE MOLECULE AND USES THEREFOR
 ; FILE REFERENCE: MP101-058P1R
 ; CURRENT APPLICATION NUMBER: US/10/122,067
 ; CURRENT FILING DATE: 2002-04-12
 ; PRIOR APPLICATION NUMBER: 60/283434
 ; PRIOR FILING DATE: 2001-04-12
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 2
 ; LENGTH: 1499
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-122-067-2

Query Match 65.4%; Score 34; DB 12; Length 1499;
 Best Local Similarity 72.7%; Pred. No. 4.2e+02;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVVPKXGMSYS 11
 Db 469 BEVVRGGGSVS 479

RESULT 12

US-10-106-698-7477
 ; Sequence 7477, Application US/10106698
 ; Publication No. US20030109690A1
 ; GENERAL INFORMATION:

; APPLICANT: Ruben et al.
 ; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
 ; FILE REFERENCE: PA005P1
 ; CURRENT APPLICATION NUMBER: US/10/106,698
 ; CURRENT FILING DATE: 2002-03-27
 ; PRIOR APPLICATION NUMBER: PCT/US00/26524
 ; PRIOR FILING DATE: 2000-09-28
 ; PRIOR APPLICATION NUMBER: US 60/157,137
 ; PRIOR FILING DATE: 1999-09-29
 ; PRIOR APPLICATION NUMBER: US 60/163,280
 ; PRIOR FILING DATE: 1999-11-03
 ; NUMBER OF SEQ ID NOS: 8564
 ; SOFTWARE: PatentIn Ver. 3.0
 ; SEQ ID NO 7477
 ; LENGTH: 99
 ; TYPE: PRT

; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: MISC_FEATURE
 ; LOCATION: (3)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; NAME/KEY: MISC_FEATURE
 ; LOCATION: (13)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; NAME/KEY: MISC_FEATURE
 ; LOCATION: (14)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; NAME/KEY: MISC_FEATURE
 ; LOCATION: (39)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; NAME/KEY: MISC_FEATURE
 ; LOCATION: (40)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; NAME/KEY: MISC_FEATURE
 ; LOCATION: (57)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; NAME/KEY: MISC_FEATURE
 ; LOCATION: (58)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; NAME/KEY: MISC_FEATURE
 ; LOCATION: (63)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; NAME/KEY: MISC_FEATURE
 ; LOCATION: (65)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; NAME/KEY: MISC_FEATURE
 ; LOCATION: (82)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; NAME/KEY: MISC_FEATURE
 ; LOCATION: (93)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 US-10-106-698-7477

Query Match 63.5%; Score 33; DB 15; Length 99;
 Best Local Similarity 66.7%; Pred. No. 32;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMSYS 11
 Db 73 LVPVGISYS 81

RESULT 13

US-10-214-766-43
 ; Sequence 43, Application US/10214766
 ; Publication No. US20030084473A1
 ; GENERAL INFORMATION:

; APPLICANT: Gocal, Greg
 ; TITLE OF INVENTION: NON-TRANSGENIC HERBICIDE RESISTANT PLANTS
 ; FILE REFERENCE: CA1138
 ; CURRENT APPLICATION NUMBER: US/10/214,766
 ; CURRENT FILING DATE: 2002-08-09
 ; PRIOR APPLICATION NUMBER: US 60/311,734
 ; PRIOR FILING DATE: 2001-08-09
 ; NUMBER OF SEQ ID NOS: 53
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 43
 ; LENGTH: 426
 ; TYPE: PRT
 ; ORGANISM: Vibrio cholerae
 US-10-214-766-43

Query Match 63.5%; Score 33; DB 15; Length 426;
 Best Local Similarity 60.0%; Pred. No. 1.6e+02;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVVPXGMSYS 10
 Db 223 EFVIPAGQSY 232

Qy 2 EVVFXGMSYS 11
|||
Db 239 EVAPAGASYN 248

Search completed: December 22, 2003, 17:32:42
Job time : 20.9333 secs

RESULT 14
US-09-92-600A-108
; Sequence 108, Application US/09992600A
; Publication No. US20030027161A1
; GENERAL INFORMATION:
; APPLICANT: Benjanin, Stephane
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 91.US4.DIV
; CURRENT APPLICATION NUMBER: US/09/992,600A
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 09/924,340
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: PCT/IB01/01715
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/302,277
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/298,698
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: JPatent
; SEQ ID NO 108
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-92-600A-108

Query Match 63.5%; Score 33; DB 11; Length 478;
Best Local Similarity 60.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 2 EVVFXGMSYS 11
|||
Db 239 EVAPAGASYN 248

RESULT 15
US-09-924-340-108
; Sequence 108, Application US/09924340
; Publication No. US20030027248A1
; GENERAL INFORMATION:
; APPLICANT: Benjanin, Stephane
; TITLE OF INVENTION: HUMAN CDNAS AND PROTEINS AND USES THEREOF
; FILE REFERENCE: 91.US2.REG
; CURRENT APPLICATION NUMBER: US/09/924,340
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/305,456
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/302,277
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/298,698
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 60/293,574
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: JPatent
; SEQ ID NO 108
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-924-340-108

Query Match 63.5%; Score 33; DB 11; Length 478;
Best Local Similarity 60.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 17:24:36 ; Search time 9.06667 Seconds
(without alignments)
116.675 Million cell updates/sec

Title: US-09-909-164-10

Perfect score: 52

Sequence: 1 BEVVPXGMSYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 76:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|--------|--------------------|
| 1 | 38 | 73.1 | 3472 | 2 | T31308 | hypothetical 367K |
| 2 | 37 | 71.2 | 840 | 2 | T39116 | probable sulfate p |
| 3 | 37 | 71.2 | 877 | 2 | T40413 | sulfate permease - |
| 4 | 36 | 69.2 | 102 | 2 | A42452 | V1 protein - tobac |
| 5 | 36 | 69.2 | 1498 | 2 | B97355 | DNA segregation At |
| 6 | 35 | 67.3 | 225 | 2 | S57810 | hypothetical prote |
| 7 | 35 | 67.3 | 425 | 2 | T24111 | hypothetical prote |
| 8 | 35 | 67.3 | 670 | 2 | S22293 | zinc finger protei |
| 9 | 35 | 67.3 | 749 | 2 | H82691 | topoisomerase IV s |
| 10 | 35 | 67.3 | 2717 | 2 | A34203 | DNA-binding protei |
| 11 | 34 | 65.4 | 156 | 2 | S54619 | hypothetical prote |
| 12 | 34 | 65.4 | 252 | 2 | H69491 | cell division inhi |
| 13 | 34 | 65.4 | 544 | 2 | C82900 | probable ABC subet |
| 14 | 33 | 63.5 | 94 | 2 | I40758 | hypothetical prote |
| 15 | 33 | 63.5 | 116 | 2 | E90544 | 50S ribosomal prot |
| 16 | 33 | 63.5 | 165 | 2 | D69493 | hypothetical prote |
| 17 | 33 | 63.5 | 253 | 2 | C81374 | hypothetical prote |
| 18 | 33 | 63.5 | 259 | 2 | T34336 | hypothetical prote |
| 19 | 33 | 63.5 | 284 | 2 | S75817 | hypothetical prote |
| 20 | 33 | 63.5 | 298 | 2 | T47670 | beta-ketoacyl-ACP |
| 21 | 33 | 63.5 | 368 | 2 | F72281 | hypothetical prote |
| 22 | 33 | 63.5 | 426 | 2 | D82163 | 3-phosphoshikimate |
| 23 | 33 | 63.5 | 466 | 2 | T43653 | cdc37 protein - fi |
| 24 | 33 | 63.5 | 653 | 2 | D82352 | iron(III) ABC tran |
| 25 | 33 | 63.5 | 890 | 2 | A30481 | bacteriocin BGN5 - |
| 26 | 33 | 63.5 | 1028 | 2 | AF3286 | ATP-dependent DNA |
| 27 | 33 | 63.5 | 1152 | 2 | D87046 | conserved hypothet |
| 28 | 33 | 63.5 | 1394 | 2 | A35626 | transforming growt |
| 29 | 33 | 63.5 | 1401 | 2 | G82336 | DNA-directed RNA p |

hypothetical prote
masking protein pr
hypothetical prote
transport protein
rho protein GDP-di
hypothetical prote
cyclin D2 - rat
cyclin D2 - mouse
cyclin D2 - human
cyclin D1 - Africa
cyclin D2 - Africa
cyclin D2 - chicke
cyclin D1 - zebra
cyclin D3 - human
cyclin D1 - human

30 33 63.5 1548 2 T04456
31 33 63.5 1712 2 A38261
32 32 61.5 84 2 E97333
33 32 61.5 175 2 P00616
34 32 61.5 223 2 T01457
35 32 61.5 279 2 B72481
36 32 61.5 288 2 JC4011
37 32 61.5 288 2 F58372
38 32 61.5 289 2 A41984
39 32 61.5 289 2 A42822
40 32 61.5 291 2 S57922
41 32 61.5 291 2 S57925
42 32 61.5 291 2 JC4579
43 32 61.5 291 2 S62730
44 32 61.5 292 2 B42822
45 32 61.5 295 2 A38977

ALIGNMENTS

RESULT 1
T31308
hypothetical 367K protein - Cenarchaeum symbiosum
C:Species: Cenarchaeum symbiosum
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 18-Feb-2000
C:Accession: T31308
R:Schleper, C.; DeLong, E.F.; Preston, C.M.; Feldman, R.A.; Wu, K.Y.; Swanson, R.V.
J. Bacteriol. 180, 5003-5009, 1998
A:Title: Genomic analysis reveals chromosomal variation in natural populations of the
A:Reference number: Z20994; MUID:98422450; PMID:9748430
A:Accession: T31308
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3472 <SCH>
A:Cross-references: EMBL:AF083072; NID:G3599393; PID:G3599394; PIDN:AAC62699.1
C:Superfamily: Cenarchaeum symbiosum hypothetical 367K protein

Query Match 73.1%; Score 38; DB 2; Length 3472;

Best Local Similarity 54.5%; Pred. No. 59;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11

DB 2294 EDVIPRGISFS 2304

RESULT 2

T39116

probable sulfate permease - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T39116

R:Hunt, C.; Aves, S.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.

submitted to the EMBL Data Library, November 1999

A:Reference number: Z21829

A:Accession: T39116

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-840 <HUN>

A:Cross-references: EMBL:ALJ32779; PIDN:CAB60015.1; GSPDB:GN00066; SPDB:SPAC869.05c

A:Experimental source: strain 972h; cosmid c869

C:Genetics:

A:Gene: SPDB:SPAC869.05c

A:Map position: 1

Query Match

Best Local Similarity 71.2%; Score 37; DB 2; Length 840;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11

DB 135 VVPQMSYA 143

```
RESULT 3
T40413
sulfate permease - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T40413
R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Jimenez Martinez, J.
submitted to the EMBL Data Library, August 1998
A:Reference number: Z21926
A:Accession: T40413
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-877 <LYN>
A:Cross-references: EMBL:AL031261; PIDN:CAA20298.1; GSPDB:GN00067; SPDB:SPBC3H7.02
A:Experimental source: strain 972h-; cosmid c3H7
C:Genetics:
A:Gene: SPDB:SPBC3H7.02
A:Map position: 2

Query Match 71.2%; Score 37; DB 2; Length 877;
Best Local Similarity 77.8%; Pred. No. 22;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
Db 148 VVPQGSYA 156

RESULT 4
A42452
V1 protein - tobacco yellow dwarf virus (strain Australia)
C:Species: tobacco yellow dwarf virus
C:Date: 15-Jan-1993 #sequence_revision 15-Jan-1993 #text_change 08-Oct-1999
C:Accession: A42452
R:Morris, B.A.M.; Richardson, K.A.; Haley, A.; Zhan, X.; Thomas, J.E.
Virology 187, 633-642, 1992
A:Title: The nucleotide sequence of the infectious cloned DNA component of tobacco yellow
A:Reference number: A42452; MUID:92188538; PMID:1546458
A:Accession: A42452
A:Molecule type: DNA
A:Residues: 1-102 <MOR>
A:Cross-references: GB:M81103; NID:g335283; PIDN:AAA47947.1; PID:g335284

Query Match 69.2%; Score 36; DB 2; Length 102;
Best Local Similarity 60.0%; Pred. No. 3.5;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVPXGMSYS 11
Db 7 QVVPGSINYS 16

RESULT 5
B97355
DNA segregation ATPase, FtsK/SpoIIIE family, YUKA B. subtilis ortholog [imported] - Clos
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: B97355
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: B97355
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1498 <KUR>
A:Cross-references: GB:AAK81629.1; PIDN:AAK81629.1; PID:g15026814; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC3709

Query Match 69.2%; Score 36; DB 2; Length 1498;
Best Local Similarity 60.0%; Pred. No. 63;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMSY 10
Db 1276 EQKIPMGMSY 1285

RESULT 6
S57810
hypothetical protein precursor (clone TPp11) - tomato
C:Species: Lycopersicon esculentum (tomato)
C:Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 21-Jul-2000
C:Accession: S57810
R:Milligan, S.B.; Gasser, C.S.
Plant Mol. Biol. 28, 691-711, 1995
A:Title: Nature and regulation of pistil-expressed genes in tomato.
A:Reference number: S57808; MUID:95375233; PMID:7647301
A:Accession: S57810
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-225 <MIL>
A:Cross-references: EMBL:U20592; NID:g924625; PIDN:AAA80497.1; PID:g924626
C:Superfamily: plant Kunitz-type proteinase inhibitor

Query Match 67.3%; Score 35; DB 2; Length 225;
Best Local Similarity 54.5%; Pred. No. 13;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMSYS 11
Db 32 DEVVPNGTYA 42

RESULT 7
T24111
hypothetical protein R10D12.10 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T24111
R:Percy, C.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19842
A:Accession: T24111
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-425 <WIL>
A:Cross-references: EMBL:Z81109; PIDN:CAB03241.1; GSPDB:GN00023; CESP:R10D12.10
A:Experimental source: clone R10D12
C:Genetics:
A:Gene: CESP:R10D12.10
A:Map position: 5
A:Introns: 23/3; 56/3; 113/3; 257/2

Query Match 67.3%; Score 35; DB 2; Length 425;
Best Local Similarity 50.0%; Pred. No. 26;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMSY 10
Db 335 EQIVPGGLQY 344

RESULT 8
S22293
zinc finger protein AT-BP2 - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 20-Sep-1999
C:Accession: S22293; I78656
R:Mitchellmore, C.; Traboni, C.; Cortese, R.
Nucleic Acids Res. 19, 141-147, 1991
```


A:Title: Isolation of two cDNAs encoding zinc finger proteins which bind to the alpha 1-
A:Reference number: I58280; MUID:91187610; PMID:1901405
A:Accession: S22293
A>Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-670 cmt>
A:Cross-references: EMBL:X54250; NID:957519; PIDN:CAA38151.1; PID:957520
A>Note: The authors did not translate the codon for residue 1
C:Superfamily: HIV-EP2 enhancer-binding protein
C:Keywords: DNA binding; transcription regulation; zinc finger

Query Match 67.3%; Score 35; DB 2; Length 670;
Best Local Similarity 66.7%; Pred. No. 43;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
||| |||
DB 376 VVPAGLTYS 384
||| |||

RESULT 9
H82691
topoisomerase IV subunit XF1353 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C:Accession: H82691
R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717; PMID:10910347
A>Note: For a complete list of authors see reference number A59328 below
A:Accession: H82691
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-749 <SIM>
A:Cross-references: GB:AE003967; GB:AE003849; NID:G9106347; PIDN:NAF84162.1; GSPDB:GN001
A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, L.E.A.; Camargo, D.M.; Carret, H
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, C.Y.; Miyaki, C.Y.
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Savasak
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF1353
C:Superfamily: DNA topoisomerase (ATP-hydrolyzing) chain A; phage T4 DNA topoisomerase

Query Match 67.3%; Score 35; DB 2; Length 749;
Best Local Similarity 77.8%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPXGMSY 10
||| |||
DB 526 EVDPGMSY 534
||| |||

RESULT 10
A34203
DNA-binding protein PRDII-BF1 - human
N:Alternate names: major histocompatibility complex enhancer-binding protein 1
C:Species: Homo sapiens (man)
C>Date: 22-Jun-1990 #sequence_revision 22-Jun-1990 #text_change 20-Sep-1999
C:Accession: A34203; A34779
R:Fan, C.M.; Maniatis, T.
Genes Dev. 4, 29-42, 1990

A:Title: A DNA-binding protein containing two widely separated zinc finger motifs that
A:Reference number: A34203; MUID:90169514; PMID:2106471
A:Accession: A34203
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2717 <PAN>
A:Cross-references: EMBL:X51435; NID:938017; PIDN:CAA35798.1; PID:938018
R:Badwin Jr., A.S.; LeClair, K.P.; Singh, H.; Sharp, P.A.
Mol. Cell. Biol. 10, 1406-1414, 1990
A:Title: A large protein containing zinc finger domains binds to related sequence ele
A:Reference number: A34779; MUID:90205817; PMID:2108316
A:Accession: A34779
A>Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 801-1072, 'N', 1074-1168, 'K', 1170-1225, 'V', 1227-1434, 'N', 1436-1607, 'I', 1609-
A:Cross-references: GB:M32019
C:Superfamily: HIV-EP2 enhancer-binding protein
C:Keywords: DNA binding; transcription regulation; zinc finger

Query Match 67.3%; Score 35; DB 2; Length 2717;
Best Local Similarity 66.7%; Pred. No. 1.9e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMSYS 11
||| |||
DB 2405 VVPAGLTYS 2413
||| |||

RESULT 11
S54619
hypothetical protein YOR013w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein O2612; hypothetical protein YOL303.3
C:Species: Saccharomyces cerevisiae
C>Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002
C:Accession: S54619; S66879
R:de Haan, M.; Maarse, A.C.; Grivell, L.A.
submitted to the EMBL Data Library, May 1995
A:Reference number: S54617
A:Accession: S54619
A:Molecule type: DNA
A:Residues: 1-156 <DEH>
A:Cross-references: EMBL:X87331; NID:gl041652; PIDN:CAA60762.1; PID:9829123
R:de Haan, M.; Grivell, L.A.; Maarse, A.C.
submitted to the Protein Sequence Database, July 1996
A:Reference number: S66877
A:Accession: S66879
A:Molecule type: DNA
A:Residues: 1-156 <DEW>
A:Cross-references: EMBL:Z74920; NID:gl420109; PIDN:CAA99201.1; PID:gl420111; MIPS:YOR
A:Experimental source: strain S288C
C:Genetics:
A:Cross-references: SGD:S0005539
A:Map position: 15R
C:Superfamily: hypothetical protein YOR013w

Query Match 65.4%; Score 34; DB 2; Length 156;
Best Local Similarity 66.7%; Pred. No. 14;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPXGMSY 10
||| |||
DB 50 EVMPGMDY 58
||| |||

RESULT 12
H69491
cell division inhibitor (minD-2) homolog - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 26-Aug-1999
C:Accession: H69491
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dods
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.
.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997
 A:Authors: Utterback, T.; Corton, M.D.; Spriggs, T.; Artisch, P.; Kaine, B.P.; Sykes, S.; Smith, H.O.; Woese, C.R.; Venter, J.C.
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon
 A:Reference number: A69250; MUID:98049343; PMID:9389475
 A:Accession: H69491
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-252 <KLE>
 A:Cross-references: GB:AB000970; GB:AE000782; NID:G2689293; PIDN:AA889318.1; PID:G264860
 C:Superfamily: cell division inhibitor mind

Query Match 65.4%; Score 34; DB 2; Length 252;
 Best Local Similarity 75.0%; Pred. No. 24;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 EVVPXGMSY 9
 ||| |||
 Db 81 EVIPAGMS 88

RESULT 13

C82900
 Probable ABC substrate-binding protein, iron UUS39 [imported] - Ureaplasma urealyticum
 C:Species: Ureaplasma urealyticum
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
 C:Accession: C82900
 R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.
 submitted to GenBank, February 2000
 A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a mir
 A:Reference number: A82870
 A:Accession: C82900
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-544 <GLA>
 A:Cross-references: GB:AB002133; GB:AF222894; NID:G6899339; PIDN:AAF30768.1; GSPDB:GN001
 A:Experimental source: serovar 3; biovar 1
 C:Genetics:
 A:Gene: ABCsbp-5; UUS39
 A:Genetic code: SGC3

Query Match 65.4%; Score 34; DB 2; Length 544;
 Best Local Similarity 70.0%; Pred. No. 55;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 BEVPXGMSY 10
 ||||| :||
 Db 135 BEVPHYLSY 144

RESULT 14

I40758
 hypothetical protein 1 - Campylobacter jejuni (fragment)
 C:Species: Campylobacter jejuni
 C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 08-Oct-1999
 C:Accession: I40758; S47317
 R:Hani, E.K.; Chan, V.L.
 J. Bacteriol. 177, 2396-2402, 1995
 A:Title: Expression and characterization of Campylobacter jejuni benzoylglycine amidohyd
 A:Reference number: I40758; MUID:95247673; PMID:7730270
 A:Accession: I40758
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-94 <RES>
 A:Cross-references: EMBL:Z36940; NID:G535805; PIDN:CAA85392.1; PID:G535806

Query Match 63.5%; Score 33; DB 2; Length 94;
 Best Local Similarity 55.6%; Pred. No. 14;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 EVVPXGMSY 10
 :||| |||
 Db 26 DIFPMSY 34

RESULT 15

E90544
 50S ribosomal protein L20 [imported] - Mycoplasma pulmonis (strain UAB CTIP)
 C:Species: Mycoplasma pulmonis
 C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
 C:Accession: E90544
 R:Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, P.; Moszer, I.
 Nucleic Acids Res. 29, 2145-2153, 2001
 A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pu
 A:Reference number: A99512; MUID:21267165; PMID:11353084
 A:Accession: E90544
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-116 <KUR>
 A:Cross-references: GB:AL445566; PID:G14089674; PIDN:CAC13434.1; GSPDB:GN00153
 A:Experimental source: strain UAB CTIP
 C:Genetics:
 A:Gene: MVPU 2610
 A:Genetic code: SGC3
 C:Superfamily: Escherichia coli ribosomal protein L20

Query Match 63.5%; Score 33; DB 2; Length 116;
 Best Local Similarity 77.8%; Pred. No. 17;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 VVPXGMSYS 11
 ||| ||| |||
 Db 68 VRPLGMSYS 76

Search completed: December 22, 2003, 17:44:57
 Job time : 9.06667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:43:51 ; Search time 4.6 Seconds
(without alignments)
112.455 Million cell updates/sec

Title: US-09-909-164-10
Perfect score: 52
Sequence: 1 EVVFXGMSYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|--------------|---------------------|
| 1 | 38 | 73.1 | 1058 | 1 CARB_FUSNN | O8qg86 fusobacteri |
| 2 | 37 | 71.2 | 877 | 1 SULH_SCHPO | Q74377 tobaccosacch |
| 3 | 36 | 69.2 | 102 | 1 Y11K TYDVA | P31619 tobacco yel |
| 4 | 36 | 69.2 | 1498 | 1 Y1A9 CLOAB | Q04351 clostridium |
| 5 | 35 | 67.3 | 2717 | 1 ZEPI HUMAN | P15822 homo sapien |
| 6 | 34 | 65.4 | 788 | 1 CY14 NEUCR | P23622 neurospora |
| 7 | 34 | 65.4 | 1499 | 1 A10C HUMAN | O60312 homo sapien |
| 8 | 33 | 63.5 | 116 | 1 RL20 MYCPU | Q98q30 mycoplasma |
| 9 | 33 | 63.5 | 165 | 1 YG49 ARCFU | O28130 archaeoglob |
| 10 | 33 | 63.5 | 253 | 1 Y990 CAMJE | P45489 campylobact |
| 11 | 33 | 63.5 | 280 | 1 CTK3 MOUSE | Q9d387 mus musculu |
| 12 | 33 | 63.5 | 426 | 1 AROA_VIBCH | Q9k8b0 vibrio chol |
| 13 | 33 | 63.5 | 466 | 1 CC37 SCHPO | O94740 schizosacch |
| 14 | 33 | 63.5 | 478 | 1 GSR2 HUMAN | Q9nzm5 homo sapien |
| 15 | 33 | 63.5 | 890 | 1 BCN5 CLOPE | P08696 clostridium |
| 16 | 33 | 63.5 | 1389 | 1 LTBS MOUSE | Q8c918 mus musculu |
| 17 | 33 | 63.5 | 1394 | 1 LTBS HUMAN | P22064 homo sapien |
| 18 | 33 | 63.5 | 1401 | 1 RPOC_VIBCH | Q9kv29 vibrio chol |
| 19 | 33 | 63.5 | 1595 | 1 LTBL HUMAN | Q14766 homo sapien |
| 20 | 33 | 63.5 | 1712 | 1 LTBL RAT | Q00918 rattus norv |
| 21 | 33 | 63.5 | 1713 | 1 LTBL MOUSE | Q8c919 mus musculu |
| 22 | 32 | 61.5 | 288 | 1 CGD2 RAT | Q04827 rattus norv |
| 23 | 32 | 61.5 | 289 | 1 CGD2 HUMAN | P30279 homo sapien |
| 24 | 32 | 61.5 | 289 | 1 CGD2 MOUSE | P30280 mus musculu |
| 25 | 32 | 61.5 | 291 | 1 CGD1_BRARE | Q90459 brachydanio |
| 26 | 32 | 61.5 | 291 | 1 CGD1_XENLA | P50755 xenopus lae |
| 27 | 32 | 61.5 | 291 | 1 CGD2_CHICK | P49706 gallus gall |
| 28 | 32 | 61.5 | 291 | 1 CGD2_XENLA | P53782 xenopus lae |
| 29 | 32 | 61.5 | 292 | 1 CGD1_CHICK | P55169 gallus gall |
| 30 | 32 | 61.5 | 292 | 1 CGD3_HUMAN | P30281 homo sapien |
| 31 | 32 | 61.5 | 295 | 1 CGD1_HUMAN | P24385 homo sapien |
| 32 | 32 | 61.5 | 295 | 1 CGD1_MOUSE | P25322 mus musculu |
| 33 | 32 | 61.5 | 295 | 1 CGD1_RAT | P39948 rattus norv |

ALIGNMENTS

RESULT 1

| ID | CARB_FUSNN | STANDARD; | PRT; | 1058 AA. |
|----|---|-----------|------|----------|
| AC | Q8RG86; | | | |
| DT | 28-FEB-2003 (Rel. 41, Created) | | | |
| DT | 28-FEB-2003 (Rel. 41, Last sequence update) | | | |
| DT | 28-FEB-2003 (Rel. 41, Last annotation update) | | | |
| DE | Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-phosphate synthetase ammonia chain). | | | |
| GN | CARB OR FMO422. | | | |
| OS | Fusobacterium nucleatum (subsp. nucleatum). | | | |
| OC | Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae; | | | |
| OC | Fusobacterium. | | | |
| OX | NCBI_TaxID=76856; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN=ATCC 25586; | | | |
| RX | MEDLINE=21886394; PubMed=11889109; | | | |
| RA | Kapatral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A., Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L., Vasileva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A., Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R., Ponstein M., Kyripides N., Overbeek R.; | | | |
| RA | "Genome sequence and analysis of the oral bacterium Fusobacterium nucleatum strain ATCC 25586."; | | | |
| RL | J. Bacteriol. 184:2005-2018(2002). | | | |
| CC | - - CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP + phosphate + L-glutamate + carbamoyl phosphate. | | | |
| CC | - - COFACTOR: Binds 3 manganese ions per subunit (By similarity). | | | |
| CC | - - PATHWAY: Arginine biosynthesis. | | | |
| CC | - - PATHWAY: Pyrimidine biosynthesis; first step. | | | |
| CC | - - SUBUNIT: Composed of two chains; the small (or glutamine) chain promotes the hydrolysis of glutamine to ammonia, which is used by the large (or ammonia) chain to synthesize carbamoyl phosphate (By similarity). | | | |
| CC | - - SIMILARITY: BELONGS TO THE CARB FAMILY. | | | |

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AB010554; AAL94625.1; ALT_INIT.

HMAP; MF 01210; -; 1.
InterPro; IPR006275; CarA_L_glu.
InterPro; IPR005483; CPhase_L.
InterPro; IPR005479; CPhase_L_D2.
InterPro; IPR005480; CPhase_L_D3.
InterPro; IPR005481; CPhase_L_N.
InterPro; IPR004362; MGS-like.
Pfam; PF0289; CPhase_L_chain; 2.
Pfam; PF02786; CPhase_L_D2; 2.
Pfam; PF02787; CPhase_L_D3; 1.

P44677 haemophilus
O75355 homo sapien
Q97616 homo sapien
P90518 crithidia f
P52384 human herpe
P32784 saccharomyc
Q64671 rattus norv
P29931 tomato aspe
P16916 escherichia
P16918 escherichia
P16917 escherichia
O95071 homo sapien

34 32 61.5 427 1 TOLB_HABIN
35 32 61.5 529 1 ENP3_HUMAN
36 32 61.5 691 1 S216_HUMAN
37 32 61.5 719 1 GSP_CRIFA
38 32 61.5 726 1 PRTF_HSV6U
39 32 61.5 759 1 SCT1_YEAST
40 32 61.5 920 1 RDD_RAT
41 32 61.5 993 1 VIA_TAV
42 32 61.5 1377 1 RHSA_ECOLI
43 32 61.5 1397 1 RHSC_ECOLI
44 32 61.5 1411 1 RHSE_ECOLI
45 32 61.5 2799 1 EDD_HUMAN

```

DR Pfam: PF02142; MGS; 1.
DR PRINTS: PR00098; CP5ASE.
DR TIGRFAMs: TIGR01369; CP5ASE1_lrg; 1.
DR PROSITE: PS00866; CP5ASE 1; 2.
DR PROSITE: PS00867; CP5ASE 2; 2.
KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
KW ATP-binding; Manganese; Complete proteome.
FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 402 546 OLIGOMERIZATION DOMAIN.
FT DOMAIN 547 929 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 930 1058 ALLOSTERIC DOMAIN.
FT REPEAT 1 546
FT REPEAT 547 1058
FT NP_BIND 153 210 ATP (POTENTIAL).
FT NP_BIND 302 352 ATP (POTENTIAL).
FT METAL 284 298 MANGANESE 1 (BY SIMILARITY).
FT METAL 298 298 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 300 300 MANGANESE 2 (BY SIMILARITY).
FT METAL 820 820 MANGANESE 3 (BY SIMILARITY).
FT METAL 832 832 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1058 AA; 117451 MW; ED7037AF77C1E39F CRC64;

Query Match 73.1%; Score 38; DB 1; Length 1058;
Best Local Similarity 60.0%; Pred. No. 6.3;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 2 EVVPXGMSYS 11
DB 190 EIVPGLNYS 199

RESULT 2
SULH SCHPO STANDARD; PRT; 877 AA.
AC 074377;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable sulfate permease C3H7.02.
GN SPBC3H7.02.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,
RA Rutherford J., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA Wellens J., Vantriel E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.-J., Xiang X., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie M.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;

```

```

RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002);
CC -1- FUNCTION: HIGH AFFINITY UPTAKE OF SULFATE INTO THE CELL (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE SLC26A FAMILY OF TRANSPORTERS.
CC -1- SIMILARITY: Contains 1 STAS domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL031261; CAA20298.1; -.
DR PIR; T40413; T40413.
DR GeneDB SPombe; SPBC3H7.02; -.
DR InterPro; IPR002645; STAS.
DR InterPro; IPR001902; Sulph_transpt.
DR Pfam; PF01740; STAS; 1.
DR Pfam; PF00916; Sulfate_transp; 1.
DR TIGRFAMs; TIGR00815; sulP; 1.
DR PROSITE; PS01130; SLC26A; 1.
DR PROSITE; PS50801; STAS; 1.
KW Transport; Transmembrane.
FT TRANSMEM 133 153 POTENTIAL.
FT TRANSMEM 161 181 POTENTIAL.
FT TRANSMEM 186 206 POTENTIAL.
FT TRANSMEM 221 241 POTENTIAL.
FT TRANSMEM 243 263 POTENTIAL.
FT TRANSMEM 292 312 POTENTIAL.
FT TRANSMEM 329 349 POTENTIAL.
FT TRANSMEM 384 404 POTENTIAL.
FT TRANSMEM 424 444 POTENTIAL.
FT TRANSMEM 461 481 POTENTIAL.
FT TRANSMEM 484 504 POTENTIAL.
FT TRANSMEM 518 538 POTENTIAL.
FT TRANSMEM 543 563 POTENTIAL.
FT DOMAIN 594 747 STAS.
SQ SEQUENCE 877 AA; 96373 MW; 56995A8493371B43 CRC64;

Query Match 71.2%; Score 37; DB 1; Length 877;
Best Local Similarity 77.8%; Pred. No. 8.4;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 3 VVPXGMSYS 11
DB 148 VVPQMSYA 156

RESULT 3
Y11K TYDVA STANDARD; PRT; 102 AA.
AC P31619;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-OCT-1993 (Rel. 27, Last annotation update)
DE Hypothetical 11.2 kDa protein.
GN V1
OS Tobacco yellow dwarf virus (strain Australia) (TYDV).
OC Viruses; ssDNA viruses; Geminiviridae; Mastrevirus.
OX NCBI_TaxID=31599;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92189538; PubMed=1546458;
RA Morris B.A.M., Richardson K.A., Haley A., Zhan X., Thomas J.E.; of
RT "The nucleotide sequence of the infectious cloned DNA component
RT tobacco yellow dwarf virus reveals features of geminiviruses
RT infecting monocotyledonous plants.";
RL Virology 187:633-642(1992).

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; M81103; AAA47947.1; --
 DR PIR; A42452; A42452.
 DR InterPro; IPR002621; Gemini mov.
 DR Pfam; PF01708; Gemini_mov; I.
 KW Hypothetical protein.
 SQ SEQUENCE 102 AA; 11178 MW; A40CEFE0AF55B67 CRC64;

Query Match 69.2%; Score 36; DB 1; Length 102;
 Best Local Similarity 60.0%; Pred. No. 1.4;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 2 EVVPKGMYS 11
 :||| :|||
 Db 7 QVPSGINYS 16

RESULT 4

Y1A9 CLOAB STANDARD; PRT; 1498 AA.
 AC Q04351;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Hypothetical protein CAC3709.
 GN CAC3709.
 OS Clostridium acetobutylicum.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1488;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 824 / DSM 792 / VCM B-1787;
 RX MEDLINE=21359325; PubMed=1146286;
 RA Noelling J., Berton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
 RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
 RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
 RA Bennett G.N., Koonin E.V., Smith D.R.;
 RT "Genome sequence and comparative analysis of the solvent-producing
 RT bacterium Clostridium acetobutylicum.";
 RL J. Bacteriol. 183:4823-4838(2001).
 RN [2]

RP SEQUENCE OF 1-108 FROM N.A.
 RC STRAIN=ATCC 824 / DSM 792 / VCM B-1787;
 RX MEDLINE=93273706; PubMed=8501044;
 RA Sauer U., Duerre P.;
 RT "Sequence and molecular characterization of a DNA region encoding a
 RT small heat shock protein of Clostridium acetobutylicum.";
 RL J. Bacteriol. 175:3394-3400(1993).
 CC -1- SIMILARITY: Contains 2 FtsK/SpoIIIE domains.
 CC -1- CAUTION: Ref.2 sequence differs from that shown due to frameshifts
 CC in positions 76 and 106.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AE007866; AA061629.1; --
 DR EMBL; X65276; CA46379.1; ALT_FRAME.
 DR PIR; B97355; B97355.
 DR InterPro; IPR002543; FtsK_SpoIIIE.
 DR Pfam; PF01580; FtsK_SpoIIIE; 2.

DR PROSITE; PS09091; FTSK_SPOIIIE; 2.
 KW Hypothetical protein; ATP-binding; Complete proteome; Repeat.
 FT DOMAIN 655 857 FTSK/SPOIIIE 1.
 FT NP_BIND 675 682 ATP (POTENTIAL).
 FT DOMAIN 1001 1188 FTSK/SPOIIIE 2.
 SQ SEQUENCE 1498 AA; 168968 MW; FF42037A335A9649 CRC64;

Query Match 69.2%; Score 36; DB 1; Length 1498;
 Best Local Similarity 60.0%; Pred. No. 24;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 BEVVPKGMYS 10
 :||| :|||
 Db 1276 EQKIPMGMSY 1285

RESULT 5

ZEPI_HUMAN STANDARD; PRT; 2717 AA.
 AC P15822;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Zinc finger protein 40 (Human immunodeficiency virus type I enhancer-
 DE binding protein 1) (HIV-EPI) (Major histocompatibility complex binding
 DE protein 1) (MBP-1) (Positive regulatory domain II binding factor 1)
 DE (PRDII-BP1).
 GN HIVBP1 OR ZNF40.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90169514; PubMed=2106471;
 RA Fan C.M., Maniatis T.;
 RT "A DNA-binding protein containing two widely separated zinc finger
 RT motifs that recognize the same DNA sequence.";
 RL Genes Dev. 4:29-42(1990).
 RN [2]
 RP STRUCTURE BY NMR OF 2113-2142.
 RX MEDLINE=91064333; PubMed=2248949;
 RA Omichinski J.G., Clore G.M., Appella E., Sakaguchi K.,
 RA Gronenborn A.M.;
 RT "High-resolution three-dimensional structure of a single zinc finger
 RT from a human enhancer binding protein in solution.";
 RL Biochemistry 29:9324-9334(1990).
 RN [3]
 RP STRUCTURE BY NMR OF 2087-2142.
 RX MEDLINE=92232684; PubMed=1567844;
 RA Omichinski J.G., Clore G.M., Robien M., Sakaguchi K., Appella E.,
 RA Gronenborn A.M.;
 RT "High-resolution solution structure of the double Cys2His2 zinc
 RT finger from the human enhancer binding protein MBP-1.";
 RL Biochemistry 31:3907-3917(1992).
 CC -1- FUNCTION: THIS PROTEIN SPECIFICALLY BINDS TO THE DNA SEQUENCE
 CC 5'-GGGCAATTC-3' WHICH IS FOUND IN THE ENHANCER ELEMENTS OF
 CC NUMEROUS VIRAL PROMOTERS SUCH AS THOSE OF SV40, CMV, OR HIV1.
 CC IN ADDITION, RELATED SEQUENCES ARE FOUND IN THE ENHANCER ELEMENTS
 CC OF A NUMBER OF CELLULAR PROMOTERS, INCLUDING THOSE OF THE CLASS I
 CC MHC, INTERLEUKIN-2 RECEPTOR, AND INTERFERON-BETA GENES. IT MAY ACT
 CC IN T-CELL ACTIVATION.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- INDUCTION: BY MITOGEN AND PHORBOL ESTER.
 CC -1- DOMAIN: CONTAINS TWO SETS OF 2 ZINC-FINGERS, WHICH ARE WIDELY
 CC SEPARATED AND RECOGNIZE THE SAME DNA SEQUENCE. THERE IS A FIFTH
 CC ZINC-FINGER IN-BETWEEN.
 CC -1- SIMILARITY: STRONG, TO HIVBP2.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 CC EMBL; X51435; CAA35798.1; -.
 CC PIR; A34203; A34203.
 CC PDB; 3ZNF; 15-JAN-92.
 CC PDB; 4ZNF; 15-JAN-92.
 CC PDB; 1BBO; 31-OCT-93.
 CC TRANSFAC; T00497; -.
 CC Genew; HGNC:4920; HIVEP1.
 CC MIM; 194540; -.
 CC GO; GO:0005634; C:nucleus; TAS.
 CC GO; GO:0003677; F:DNA binding activity; TAS.
 CC InterPro; IPR007087; Znf_C2H2.
 CC Pfam; PF00096; zf-C2H2; 5.
 CC SMART; SMO0355; Znf_C2H2; 4.
 CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
 CC PROSITE; PS0157; ZINC_FINGER_C2H2_2; 4.
 CC Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
 KW Nuclear protein; Repeat; 3D-structure.
 FT ZN_FING 406 428
 FT ZN_FING 434 456
 FT ZN_FING 958 981
 FT ZN_FING 2087 2109
 FT ZN_FING 2115 2139
 FT DOMAIN 803 806
 FT STRAND 2088 2088
 FT TURN 2090 2092
 FT STRAND 2095 2095
 FT HELIX 2099 2108
 FT TURN 2109 2109
 FT STRAND 2115 2116
 FT STRAND 2123 2124
 FT HELIX 2127 2135
 SQ SEQUENCE 2717 AA; 297217 MW; D45D3CA951FEA561 CRC64;
 Query Match 67.3%; Score 35; DB 1; Length 2717;
 Best Local Similarity 66.7%; Pred. No. 73;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 Oy 3 VVPXGMSYS 11
 Db 2405 VVPAGLTYS 2413
 RESULT 6
 CY14 NEUCR
 ID CY14 NEUCR STANDARD; PRT; 788 AA.
 AC P23622;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Sulfate permease II.
 GN CYS-14.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
 OX NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91129256; PubMed=1825178;
 RA Ketter J.S., Jarai G., Fu Y.-H., Marzluf G.A.;
 RT "Nucleotide sequence, messenger RNA stability, and DNA recognition
 RT elements of cys-14, the structural gene for sulfate permease II in
 RT Neurospora crassa."
 RL Biochemistry 30:1780-1787 (1991).
 RN [2]
 RP PROBABLE REVISIONS.
 RX MEDLINE=94188926; PubMed=8140616;
 RA Sandal N.N., Marcker K.A.;
 RT "Similarities between a soybean nodulin, Neurospora crassa sulphate
 RT permease II and a putative human tumour suppressor.";

Trends Biochem. Sci. 19:19-19 (1994).
 RL -1- FUNCTION: UPTAKE OF SULFATE INTO THE CELL.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- INDUCTION: Highly expressed, but only in cells subject to sulfur
 CC limitation, and it is turned on by the positive-acting Cys-3
 CC sulfur regulatory protein.
 CC -1- MISCELLANEOUS: SULFATE PERMEASE II IS MAINLY FOUND IN MYCELIA.
 CC -1- SIMILARITY: BELONGS TO THE SLC36A FAMILY OF TRANSPORTERS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL; M59167; AAA33615.1; ALT SEQ.
 DR InterPro; IPR001902; Sulph transp.
 DR Pfam; PF00916; Sulfate transp; 1.
 DR TIGRPFams; TIGR00815; sulP; 1.
 DR PROSITE; PS01130; SLC26A; 1.
 DR Transmembrane; Glycoprotein.
 FT TRANSMEM 71 91
 FT TRANSMEM 103 123
 FT TRANSMEM 128 148
 FT TRANSMEM 171 191
 FT TRANSMEM 193 213
 FT TRANSMEM 271 291
 FT TRANSMEM 326 346
 FT TRANSMEM 363 383
 FT TRANSMEM 451 471
 FT TRANSMEM 474 494
 FT CARBOHYD 23 23
 FT CARBOHYD 578 578
 SQ SEQUENCE 788 AA; 87864 MW; 4FC604B60798CE77 CRC64;
 Query Match 65.4%; Score 34; DB 1; Length 788;
 Best Local Similarity 66.7%; Pred. No. 32;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 Oy 3 VVPXGMSYS 11
 Db 90 VVPQGMAYA 98

RESULT 7
 A10C HUMAN
 ID A10C HUMAN STANDARD; PRT; 1499 AA.
 AC O60312; O969T4;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Potential phospholipid-transporting ATPase VC (EC 3.6.3.1) (ATPVC)
 DB (Aminophospholipid translocase VC).
 GN ATP10C OR ATPVC OR KIAA0566.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21252579; PubMed=11326269;
 RA Meguro M., Kashiwagi A., Mitsuya K., Nakao M., Kondo I., Saitoh S.,
 RA Oshimura M.;
 RT "A novel maternally expressed gene, ATP10C, encodes a putative
 RT aminophospholipid translocase associated with Angelman syndrome."
 RL Nat. Genet. 28:19-20 (2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21313119; PubMed=11353404;
 RA Herzog L.B.K., Kim S.-J., Cook E.H. Jr., Ledbetter D.H.;
 RT "The human aminophospholipid-transporting ATPase gene ATP10C maps

adjacent to UBE3A and exhibits similar imprinted expression.";
 [3]
 RN SEQUENCE OF 337-1499 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=98290545; PubMed=9628591;
 RA Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
 RA Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. IX.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro.";
 RL DNA Res. 5:31-39(1998).
 CC -|- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.
 CC -|- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
 CC -|- TISSUE SPECIFICITY: Wide expression, with highest levels in
 CC kidney, followed by lung, brain, prostate, testis, ovary, and
 CC small intestine.
 CC -|- DISEASE: Defects in ATP10C are a cause of Angelman syndrome (AS)
 CC [MIM:105830]; also known as 'happy puppet syndrome'. It is
 CC characterized by features of severe motor and intellectual
 CC retardation, microcephaly, ataxia, frequent jerky limb movements
 CC and flapping of the arms and hands, hypotonia, hyperactivity,
 CC hypoglycemia, seizures, absence of speech, frequent smiling
 CC and episodes of paroxysmal laughter, and an unusual facies
 CC characterized by macrostomia, a large mandible and open-mouthed
 CC expression, a great propensity for protruding the tongue ('tongue
 CC thrusting'), and an occipital groove.
 CC -|- SIMILARITY: Belongs to the cation transport ATPases family (P-type
 CC ATPases). Subfamily IV.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AB051358; BAB47392.1; -;
 DR EMBL; AY029504; AAK33100.1; -;
 DR EMBL; AY029487; AAK33100.1; JOINED.
 DR EMBL; AY029488; AAK33100.1; JOINED.
 DR EMBL; AY029489; AAK33100.1; JOINED.
 DR EMBL; AY029490; AAK33100.1; JOINED.
 DR EMBL; AY029491; AAK33100.1; JOINED.
 DR EMBL; AY029492; AAK33100.1; JOINED.
 DR EMBL; AY029493; AAK33100.1; JOINED.
 DR EMBL; AY029494; AAK33100.1; JOINED.
 DR EMBL; AY029495; AAK33100.1; JOINED.
 DR EMBL; AY029496; AAK33100.1; JOINED.
 DR EMBL; AY029497; AAK33100.1; JOINED.
 DR EMBL; AY029498; AAK33100.1; JOINED.
 DR EMBL; AY029499; AAK33100.1; JOINED.
 DR EMBL; AY029500; AAK33100.1; JOINED.
 DR EMBL; AY029501; AAK33100.1; JOINED.
 DR EMBL; AY029502; AAK33100.1; JOINED.
 DR EMBL; AY029503; AAK33100.1; JOINED.
 DR EMBL; AB011138; BAA25492.1; -;
 DR Genbank; HGNC:13547; ATP10C.
 DR MIM; 605855; -;
 DR MIM; 105830; -;
 DR GO; GO:0016021; C:integral to membrane; NAS.
 DR GO; GO:0004012; F:phospholipid-translocating ATPase activity; NAS.
 DR GO; GO:0008360; F:regulation of cell shape; NAS.
 DR InterPro; IPR001757; ATPase E1-E2.
 DR InterPro; IPR006539; Flippase.
 DR InterPro; IPR005834; Hydrolase.
 DR Pfam; PF00702; Hydrolase.1.
 DR PRINTS; PR00119; CATATPASE.
 DR TIGRFAMs; TIGR01652; ATPase-Flipid; 1.
 DR TIGRFAMs; TIGR01494; ATPase P-type; 6.
 DR PROSITE; PS00154; ATPASE E1 E2; 1.
 DR Hydrolase; Transmembrane; Phosphorylation; Magnesium; ATP-binding;

Multigene family.
 KW DOMAIN 1 86
 FT TRANSMEM 87 106
 FT DOMAIN 107 110
 FT TRANSMEM 111 128
 FT DOMAIN 129 302
 FT TRANSMEM 310 332
 FT DOMAIN 333 362
 FT TRANSMEM 363 384
 FT DOMAIN 385 1087
 FT TRANSMEM 1088 1108
 FT DOMAIN 1109 1119
 FT TRANSMEM 1120 1140
 FT DOMAIN 1141 1170
 FT TRANSMEM 1171 1192
 FT DOMAIN 1193 1199
 FT TRANSMEM 1200 1222
 FT DOMAIN 1223 1228
 FT TRANSMEM 1229 1249
 FT DOMAIN 1250 1267
 FT TRANSMEM 1268 1292
 FT DOMAIN 1293 1499
 FT MOD RES 427 427
 FT METAL 1031 1031
 FT METAL 1035 1035
 FT DOMAIN 467 470
 FT CONFLICT 388 388
 SQ SEQUENCE 1499 AA; 167687 MW; D4996A4D0635A68D CRC64;
 Query Match 65.4%; Score 34; DB 1; Length 1499;
 Best Local Similarity 72.7%; Pred. No. 64;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Qy 1 BEVVPXGMSVS 11
 Db 469 BEVVPXGMSVS 479
 RESULT 8
 RL20_MYCPU STANDARD; PRT; 116 AA.
 AC Q98QV0;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 508 ribosomal protein L20.
 OS Mycoplasma pulmonis.
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2107;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=UAB CTIP;
 RX MEDLINE=21267165; PubMed=11353084;
 RA Chambaud I., Heilig R., Ferris S., Barbe V., Samsen D., Galleson P.,
 RA Moszer I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,
 RA Blanchard A.;
 RT "The complete genome sequence of the murine respiratory pathogen
 RT Mycoplasma pulmonis.";
 RL Nucleic Acids Res. 29:2145-2153(2001).
 CC -|- FUNCTION: This protein binds directly to 23s ribosomal RNA and is
 CC necessary for the in vitro assembly process of the 50s ribosomal
 CC subunit. It is not involved in the protein synthesizing functions
 CC of that subunit (By similarity).
 CC -|- SIMILARITY: BELONGS TO THE L20P FAMILY OF RIBOSOMAL PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC


```

CC EMBL; AL445563; CAC13434.1; --
DR PIR; E90544; E90544.
DR MYPLIST; MYPL 2610; --
DR HAMAP; MF_00382; -- 1.
DR InterPro; IPR005813; L20.
DR Pfam; PF00453; Ribosomal_L20; 1.
DR PRINTS; PR00062; RibosomalL20.
DR PRODOM; PD002389; L20; 1.
DR TIGRFAMs; TIGR01032; rplT_bact; 1.
DR PROSITE; PS00937; RIBOSOMAL_L20; 1.
KW Ribosomal protein; rRNA-binding; Complete proteome.
SQ SEQUENCE 116 AA; 13565 MW; C59C748901B18F14 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 116;
Best Local Similarity 77.8%; Pred. No. 7.1;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 3 VVPGXGMSYS 11
Db 68 VRPLGMSYS 76

RESULT 9
YJ49 ARCFU
ID YJ49 ARCFU STANDARD; PRT; 165 AA.
AC O28330;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Hypothetical protein AF1949.
GN AF1949.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA Sadow P.M., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

DR EMBL; AE000968; AAB89307.1; --
DR PIR; D69493; D69493.
DR TIGR; AF1949; --
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 7 27 POTENTIAL.
FT TRANSMEM 141 161 POTENTIAL.
SQ SEQUENCE 165 AA; 17588 MW; B8C17054810ADBFB CRC64;

Query Match 63.5%; Score 33; DB 1; Length 165;

```

```

Best Local Similarity 60.0%; Pred. No. 10;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 1 EVVPXGMSY 10
Db 60 EESIPDGASY 69

RESULT 10
Y990 CAMJE
ID Y990 CAMJE STANDARD; PRT; 253 AA.
AC P45489; Q9PNV0;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein Cj0990c.
GN Cj0990c.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Kettle J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA Jørgensen K., Kariyasek A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrall B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences.";
RL Nature 403:665-668(2000).
RN [2]
RP SEQUENCE OF 160-253 FROM N.A.
RC STRAIN=ATCC 43431 / TGH 9011;
RX MEDLINE=95247673; PubMed=7730270;
RA Hani B.K., Chan V.L.;
RT "Expression and characterization of Campylobacter jejuni
RT benzoylglycine amidohydrolase (Hippuricase) gene in Escherichia
RT coli.";
RL J. Bacteriol. 177:2396-2402(1995).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

DR EMBL; AL139076; CAB73246.1; --
DR EMBL; Z36940; CAA85392.1; --
DR PIR; C81374; C81374.
DR PIR; I40758; I40758.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 253 AA; 29783 MW; F56D3FF3265F8A6A CRC64;

Query Match 63.5%; Score 33; DB 1; Length 253;
Best Local Similarity 55.6%; Pred. No. 16;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Oy 2 EVVPXGMSY 10
Db 185 DIPFGMSY 193

RESULT 11
CTX3 MOUSE
ID CTX3 MOUSE STANDARD; PRT; 280 AA.
AC Q9D387; Q9CXQ4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)

```


15-SEP-2003 (Rel. 42, Last annotation update)
DR DE GN OS OC CX OK RN
Mus musculus (Mouse)
Eukaryota; Metazoa;
Mammalia; Eutheria;
NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC STRAIN=C57BL/6J; TISSUE=Embryonic head;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Azawa K., Ozawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Stauber F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayaehizaki Y.,
RA "Functional annotation of a full-length mouse cDNA collection."
Nature 409:685-690(2001).
[2]
SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Mammary fibroblast;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.K., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McWain P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S.E., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villallon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalek U., Smalish D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -[- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC -[- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q9D387-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9D387-2; Sequence=VSP 003820;
CC -[- CAUTION: Ref.1 sequence differs from that shown due to frameshifts
CC in positions 174 and 239.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

EMBL; AK014127; BAB29169.1; -
EMBL; AK018222; BAB31124.1; ALT_FRAME.
EMBL; BC004791; AAH04791.1; -
MGD; MGI:1920368; 3110035N03Rik.
MGD; MGI:1923411; 6330527O06Rik.
Transmembrane; Signal; Alternative splicing.
FT SIGNAL 1 29 POTENTIAL.
FT CHAIN 30 280 PROTEIN C20ORF103 HOMOLOG.
FT DOMAIN 30 235 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 236 256 POTENTIAL.
FT DOMAIN 257 280 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 35 35 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 53 53 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 102 102 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPPLIC 1 118 Missing (in isoform 2).
FT FTid=VSP 003820
FT CONFLICT 221 221 E -> V (IN REF. 1; BAB31124).
FT CONFLICT 230 230 Q -> P (IN REF. 1; BAB31124).
FT CONFLICT 238 238 P -> A (IN REF. 1; BAB31124).
SQ SEQUENCE 280 AA; 31721 MW; FA11D7BF9D5CCBP CRC64;
Query Match 63.5%; Score 33; DB 1; Length 280;
Best Local Similarity 75.0%; Pred. No. 18;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 3 VVPGXMSY 10
DB 173 VTPAGMSY 180
RESULT 12
ID AROA VIBCH STANDARD; PRT; 426 AA.
AC Q9K3E0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19) (5-
DE enolpyruvylshikimate-3-phosphate synthase) (BPS synthase) (BPSPS).
GN AROA OR VC1732.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA Mcdonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae".
RL Nature 406:477-483(2000).
CC -[- CATALYTIC ACTIVITY: Phosphoenolpyruvate + 3-phosphoshikimate =
CC phosphate + 5-O-(1-carboxyvinyl)-3-phosphoshikimate.
CC -[- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
CC sixth step.
CC -[- SUBUNIT: Monomer (By similarity).
CC -[- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -[- SIMILARITY: Belongs to the BPS synthase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

or send an email to license@isb-sib.ch).
CC EMBL; AE004251; AAF94882.1; -.
CC PIR; D82163; D82163.
CC TIGR; VC1732; -.
CC HAMAP; MF 00210; -.
CC InterPro; IPR006264; AroA.
CC InterPro; IPR001986; EPSP synthase.
CC Pfam; PF00275; EPSP synthase; 1.
CC ProDom; PD001867; EPSP synthase; 1.
CC TIGRFAMs; TIGR01356; aroA; 1.
CC PROSITE; PS00104; EPSP SYNTHASE_1; 1.
CC PROSITE; PS00885; EPSP SYNTHASE_2; 1.
KW Aromatic amino acid biosynthesis; Transferase; Complete proteome.
SQ SEQUENCE 426 AA; 46101 MW; 38852D6483BPE1C3 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 426;
Best Local Similarity 60.0%; Pred. No. 28;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVVPXGMSY 10
Db 223 EFVIPAGQSY 232

RESULT 13
CC37_SCHPO
ID CC37_SCHPO STANDARD; PRT; 466 AA.
AC 094740;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hsp90 co-chaperone Cdc37 (Hsp90 chaperone protein kinase-targeting
DE subunit) (Cell division control protein 37).
GN CDC37 OR SPBC9B6.10.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
[1]
RP SEQUENCE FROM N.A.
RA Westwood P.K., Preston N.C., Fantes P.A.;
RA "Schizosaccharomyces pombe cdc37 gene.";
RA Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA Jones K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grynoprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie M.W., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;

```

```

"The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -I- FUNCTION: With Hsp90 it forms a complex that binds to several
CC kinases, resulting in stabilization and promotion of their
CC activity (By similarity).
CC -I- SUBUNIT: Forms a complex with Hsp90. Interacts with a number of
CC kinases (By similarity).
CC -I- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -I- SIMILARITY: BELONGS TO THE CDC37 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ132377; CAB38758.1; -.
CC EMBL; AJ132376; CAB38757.1; -.
CC EMBL; AL049769; CAB42371.2; -.
CC PIR; T43653; T43653.
CC GeneDB_Spombe; SPBC9B6.10; -.
CC InterPro; IPR004918; Cdc37.
CC Pfam; PF03234; Cdc37; 1.
KW Chaperone; Cell division; Cell cycle.
SQ SEQUENCE 466 AA; 52554 MW; 647238B34CABB3C5 CRC64;

Query Match 63.5%; Score 33; DB 1; Length 466;
Best Local Similarity 50.0%; Pred. No. 30;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVVPXGMSY 10
Db 98 DSAIPGMSY 107

RESULT 14
GSR2_HUMAN
ID GSR2_HUMAN STANDARD; PRT; 478 AA.
AC Q9NZM5; Q9BTC6; Q9HAX6; Q9NPPI; Q9NPFR4; Q9UF12;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Glioma tumor suppressor candidate region gene 2 protein (p60).
GN GLTSCR2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A. AND TISSUE SPECIFICITY.
RX MEDLINE=20175430; PubMed=10708517;
RA Smith J.S., Tachibana I., Pohl U., Lee H.K., Thanarajasingam U.,
RA Portier B.P., Ueki K., Billings S., Ramasamy S., Mohrenweiser H.W.,
RA Scheithauer B.W., Louie D.N., Jenkins R.B.;
RA "A transcript map of the chromosome 19q-Arm glioma tumor suppressor
RA region.";
RT region.;
RN Genomics 64:44-50(2000).
[2]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh F.,
RA Diatchenko L., Narusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., McQuellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

```

RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RA Schmerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [3]
RP SEQUENCE OF 9-478 FROM N.A., AND SUBCELLULAR LOCATION.
RX MEDLINE=99214318; PubMed=10196275;
RA Bruni R., Fineschi B., Ogle W.O., Roizman B.;
RT "A novel cellular protein, p60, interacting with both herpes simplex
virus 1 regulatory proteins ICP22 and ICP0 is modified in a
cell-type-specific manner and is recruited to the nucleus after
infection.";
RL J. Virol. 73:3810-3817 (1999).
RN [4]
RP SEQUENCE OF 12-478 FROM N.A.
RA Andreu N., Estivill X., Escarceller M., Sumoy L.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 218-477 FROM N.A.
RC TISSUE-Testis;
RA Poustka A., Klein M., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
CC 1- SUBUNIT: Interacts with HSV-1 early proteins ICP22 and ICP0.
CC 1- SUBCELLULAR LOCATION: Nuclear.
CC 1- TISSUE SPECIFICITY: Expressed at high levels in heart and
pancreas, moderate levels in placenta, liver, skeletal muscle, and
kidney, and low levels in brain and lung.
CC 1- SIMILARITY: BELONGS TO THE GLTSCR2 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
DR EMBL; AF182076; AAF62873.1; -
DR EMBL; BC004229; AAH04229.1; -
DR EMBL; BC006311; AAH06311.1; -
DR EMBL; BC010095; AAH10095.1; -
DR EMBL; AF296124; AAG30413.1; -
DR EMBL; AF359335; CAB94786.1; -
DR EMBL; AL359336; CAB94787.1; -
DR EMBL; AL122063; CAB59242.1; -
DR SWISS-2DPAGE; Q9NZMS; HUMAN.
DR Genew; HGNC:4333; GLTSCR2.
DR MIM; 605691; -
DR GO; GO:0008181; F: tumor suppressor; TAS.
KW Nuclear protein; Polymorphism.
FT VARIANT 389 389
FT R -> Q.
FT /FTID=VAR_011486.
FT GGS -> HEG (IN REF. 2; AAH04229).
FT G -> R (IN REF. 3).
FT CONFLICT 4 6
FT CONFLICT 9 9
FT CONFLICT 146 191
FT RRKEQLWEKLAKQGLPREVRRQAARLNPSPATRAKPGQPD
FT TVERP -> SGKSSYGRSWPGRASSPGCAQGPSVPAQPCFN
FT KGNPAPGHRTAA (IN REF. 3).
FT SDNFLDRPLVQDSFFLE -> LNNPKFVVPWPGCLPFG
FT (IN REF. 3).
FT A -> S (IN REF. 2; AAH04229).
FT D -> H (IN REF. 3).
FT PEGNLRDRFKRSFORNMIEPRERAKRKYKVKLVKRAFR
FT REIQ -> VLVTCGEGAPCVMTSLPLVPRGVRHGGCP
FT WAGPVGPMRPG (IN REF. 5).
FT EGNLRDRFKRSFORNMIEPRERAKRKYKVKLVKRAFR
FT BIQL -> RGQHSFETSGRAFRGI (IN REF. 3).
FT CONFLICT 434 478
FT SEQUENCE 478 AA; 54417 MW; 7F18923E348CB52B CRC64;

Query Match 63.5%; Score 33; DB 1; Length 478;
Best Local Similarity 60.0%; Pred. No. 31;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 2 EVVFXGMSYS 11
Db 239 EVAPAGASYN 248

RESULT 15
BCNS_CLOPE
ID BCNS_CLOPE STANDARD; PRT; 890 AA.
AC P08696;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE Bacteriocin BCNS.
GN BGN.
OS Clostridium perfringens.
OG Plasmid pIP404.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CPN50;
RX MEDLINE=88336297; PubMed=2901768;
RA Garnier T., Cole S.T.;
RT "Complete nucleotide sequence and genetic organization of the
bacteriocinogenic plasmid, pIP404, from Clostridium perfringens.";
RL Plasmid 19:134-150 (1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CPN50;
RX MEDLINE=87057020; PubMed=2877971;
RA Garnier T., Cole S.T.;
RT "Characterization of a bacteriocinogenic plasmid from Clostridium
perfringens and molecular genetic analysis of the
bacteriocin-encoding gene.";
RL J. Bacteriol. 168:1189-1196 (1986).
RN [3]
RP SEQUENCE OF 1-14 FROM N.A.
RC STRAIN=CPN50;
RX MEDLINE=89039249; PubMed=2460717;
RA Garnier T., Cole S.T.;
RT "Studies of UV-inducible promoters from Clostridium perfringens in
vivo and in vitro.";
RL Mol. Microbiol. 2:607-614 (1988).
CC 1- FUNCTION: MAY FUNCTION AS AN IONOPHORE.
CC 1- INDUCTION: By UV irradiation.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
DR EMBL; M14481; AAA98248.1; -
DR EMBL; M32882; AAA98249.1; -
DR PIR; A30481; A30481.
DR InterPro; IPR003646; SH3_bac.
DR InterPro; IPR000834; Zn_carbopept.
DR Pfam; PF00246; Zn_carbopept; 1.
DR SMART; SM00287; SH3b; 3.
KW Antibiotic; Bacteriocin; Plasmid.
FT DOMAIN 815 869 HYDROPHOBIC.
FT SEQUENCE 890 AA; 96699 MW; F4E5E8971C31C6C6 CRC64;
Query Match 63.5%; Score 33; DB 1; Length 890;
Best Local Similarity 66.7%; Pred. No. 60;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 EVVPXGMSY 10

|||||:

Db 170 EVVPGGFTY 178

Search completed: December 22, 2003, 17:42:26
Job time : 4.6 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 17:27:26 ; Search time 25.2 Seconds
(without alignments)
112.642 Million cell updates/sec

Title: US-09-909-164-10
Perfect score: 52
Sequence: 1 EEVVPXGMSYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_23.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_virus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|-----------|--------------------|
| 1 | 39 | 75.0 | 1044 | 16 Q8DIH0 | Q8DIH0 synechococ |
| 2 | 38 | 73.1 | 387 | 16 Q98FX1 | Q98FX1 rhizobium 1 |
| 3 | 38 | 73.1 | 3472 | 1 O74056 | O74056 cenarchaeum |
| 4 | 37 | 71.2 | 840 | 3 Q9URY8 | Q9URY8 schizosacch |
| 5 | 37 | 71.2 | 1123 | 16 Q8WD4 | Q8WD4 mycoplasma |
| 6 | 36 | 69.2 | 471 | 11 Q8R126 | Q8R126 mus musculu |
| 7 | 36 | 69.2 | 484 | 11 Q8VD18 | Q8VD18 mus musculu |
| 8 | 36 | 69.2 | 484 | 11 Q8BTX4 | Q8BTX4 mus musculu |
| 9 | 36 | 69.2 | 484 | 11 Q8BK35 | Q8BK35 mus musculu |
| 10 | 35 | 67.3 | 225 | 10 Q40129 | Q40129 lycopersico |
| 11 | 35 | 67.3 | 425 | 5 Q9XVK4 | Q9XVK4 caenorhabdi |
| 12 | 35 | 67.3 | 495 | 11 Q8CID7 | Q8CID7 mus musculu |
| 13 | 35 | 67.3 | 556 | 4 Q43733 | Q43733 homo sapien |
| 14 | 35 | 67.3 | 583 | 5 Q9BH83 | Q9BH83 plasmodium |
| 15 | 35 | 67.3 | 583 | 5 Q9BHAS | Q9BHAS plasmodium |
| 16 | 35 | 67.3 | 583 | 5 Q815S7 | Q815S7 plasmodium |

| | | | | | |
|----|----|------|------|-----------|--------------------|
| 17 | 35 | 67.3 | 670 | 11 Q01487 | Q01487 rattus ratt |
| 18 | 35 | 67.3 | 747 | 16 Q8PM16 | Q8PM16 xanthomonas |
| 19 | 35 | 67.3 | 747 | 16 Q8PAT2 | Q8PAT2 xanthomonas |
| 20 | 35 | 67.3 | 749 | 16 Q9PDM6 | Q9PDM6 xylella fas |
| 21 | 35 | 67.3 | 1902 | 4 Q14122 | Q14122 homo sapien |
| 22 | 34 | 65.4 | 156 | 3 Q12479 | Q12479 saccharomyc |
| 23 | 34 | 65.4 | 219 | 17 Q971S2 | Q971S2 sulfolobus |
| 24 | 34 | 65.4 | 252 | 17 Q28342 | Q28342 archaeoglob |
| 25 | 34 | 65.4 | 290 | 4 Q96MU1 | Q96MU1 homo sapien |
| 26 | 34 | 65.4 | 387 | 16 Q92MD6 | Q92MD6 rhizobium m |
| 27 | 34 | 65.4 | 489 | 4 Q81YN3 | Q81YN3 homo sapien |
| 28 | 34 | 65.4 | 541 | 16 Q98BP5 | Q98BP5 rhizobium 1 |
| 29 | 34 | 65.4 | 544 | 16 Q9PDQ2 | Q9PDQ2 ureaplasma |
| 30 | 34 | 65.4 | 842 | 3 Q9URR4 | Q9URR4 penicillium |
| 31 | 34 | 65.4 | 899 | 16 Q8G415 | Q8G415 bifidobacte |
| 32 | 34 | 65.4 | 1049 | 16 Q8XT05 | Q8XT05 ralstonia s |
| 33 | 33 | 63.5 | 143 | 17 Q8TX62 | Q8TX62 methanopyru |
| 34 | 33 | 63.5 | 166 | 16 Q8PPF5 | Q8PPF5 xanthomonas |
| 35 | 33 | 63.5 | 193 | 2 Q8VUA8 | Q8VUA8 lactococcus |
| 36 | 33 | 63.5 | 208 | 2 Q8KTQ4 | Q8KTQ4 candidatus |
| 37 | 33 | 63.5 | 209 | 16 Q8RE56 | Q8RE56 fusobacteri |
| 38 | 33 | 63.5 | 284 | 16 P74187 | P74187 synechocyst |
| 39 | 33 | 63.5 | 298 | 10 Q9M3C0 | Q9M3C0 arabidopsis |
| 40 | 33 | 63.5 | 326 | 12 Q9QCE7 | Q9QCE7 soil-borne |
| 41 | 33 | 63.5 | 326 | 12 Q9Q9Q7 | Q9Q9Q7 soil-borne |
| 42 | 33 | 63.5 | 326 | 12 Q9DJG4 | Q9DJG4 soil-borne |
| 43 | 33 | 63.5 | 326 | 12 Q9Q9Q9 | Q9Q9Q9 soil-borne |
| 44 | 33 | 63.5 | 326 | 12 Q91DN1 | Q91DN1 soil-borne |
| 45 | 33 | 63.5 | 326 | 12 Q9Q9Q5 | Q9Q9Q5 soil-borne |

ALIGNMENTS

RESULT 1

Q8DIH0 ID Q8DIH0 PRELIMINARY; PRT; 1044 AA.
AC Q8DIH0
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DB Multidrug efflux transporter.
GN TLL1618.
OS Synechococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=32046;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-1;
RX MEDLINE=22225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Kato H., Sasamoto S.,
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA Shimpo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
Thermosynechococcus elongatus BP-1";
EL DNA Res. 9:123-130(2002).
DR EMBL; AP005374; BAC09170.1; -.
KW Complete proteome.
SQ SEQUENCE 1044 AA; 113205 MW; 00E9C13F0F636D2F CRC64;

Query Match 75.0%; Score 39; DB 16; Length 1044;
Best Local Similarity 63.6%; Pred. No. 24;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGMSYS 11
|||:|:|:
DB 843 EEVLPNGIGYS 853

RESULT 2

Q98FX1 ID Q98FX1 PRELIMINARY; PRT; 387 AA.

AC Q98FX1; 01-OCT-2001 (TReMBLrel. 18, Created)
 DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
 DT 01-OCT-2001 (TReMBLrel. 18, Last annotation update)
 DE Hippurate hydrolase.
 GN MLR3583.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Phyllobacteriaceae; Mesorhizobium.
 OX NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Ideawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 RT Mesorhizobium loti-";
 RL DNA Res. 7:331-338(2000).
 DR EMBL; AP003002; BAB50445.1; -;
 DR InterPro; IPR002933; Peptidase M20.
 DR Pfam; PF01546; Peptidase M20; 1.
 KW Hydrolase; Complete proteome.
 SQ SEQUENCE 387 AA; 41180 MW; 131BFF8E64306829 CRC64;

Query Match 73.1%; Score 38; DB 16; Length 387;
 Best Local Similarity 60.0%; Pred. No. 13;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Oy 1 BEVVPXGMSYS 10
 Db 367 DEAIIPHGMSY 376
 :|:|:|:|:|

RESULT 3
 O74056 PRELIMINARY; PRT; 3472 AA.
 ID O74056
 AC O74056;
 DT 01-NOV-1998 (TReMBLrel. 08, Created)
 DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE Hypothetical 367.1 kDa protein.
 OS Cenarchaeum symbiosum.
 OC Archaea; Crenarchaeota; Thermoprotei; Cenarchaeales; Cenarchaeaceae;
 OC Cenarchaeum.
 OX NCBI_TaxID=46770;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=B;
 RX MEDLINE=98422450; PubMed=9748430;
 RA Schleper C., DeLong E.F., Preston C.M., Feldman R.A., Wu K.Y.,
 RA Swanson R.V.;
 RT "Genomic analysis reveals chromosomal variation in natural populations
 RT of the uncultured psychrophilic archaeon Cenarchaeum symbiosum.";
 RL J. Bacteriol. 180:5003-5009(1998).
 DR EMBL; AF083072; AACG2699.1; -;
 DR InterPro; IPR000515; BPD transp.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF00400; WD40; 4.
 DR SMART; SM00320; WD40; 2.
 DR PROSITE; PS00402; BPD_TRANSF_INN_MEMBER; 1.
 KW Hypothetical protein; Repeat; WD repeat.
 SQ SEQUENCE 3472 AA; 367058 MW; 37F80707030F9355 CRC64;

Query Match 73.1%; Score 38; DB 1; Length 3472;
 Best Local Similarity 54.5%; Pred. No. 1.5e+02;
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Oy 1 BEVVPXGMSYS 11
 Db 367 DEAIIPHGMSY 376
 :|:|:|:|:|

Db 2294 EDVIPRGISFS 2304
 RESULT 4
 Q9URY8 PRELIMINARY; PRT; 840 AA.
 ID Q9URY8
 AC Q9URY8;
 DT 01-MAY-2000 (TReMBLrel. 13, Created)
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE Probable sulfate permease.
 GN SPAC869.05C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972h-;
 RA Hunt C., Aves S., McDougall R.C., Rajandream M.A., Barrell B.G.;
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL132779; CAB60015.1; -;
 DR GeneDB; Spombe; SPAC869.05c; -;
 DR InterPro; IPR002645; STAS.
 DR InterPro; IPR001902; Sulph_transpt.
 DR Pfam; PF01740; STAS; 1.
 DR Pfam; PF00916; Sulfate_transp; 1.
 DR TIGRFAMs; TIGR00815; sulP; 1.
 DR PROSITE; PS0801; STAS; 1.
 SQ SEQUENCE 840 AA; 93517 MW; ED4833B162B69077 CRC64;

Query Match 71.2%; Score 37; DB 3; Length 840;
 Best Local Similarity 77.8%; Pred. No. 51;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 3 VVPXGMSYS 11
 Db 135 VVPQGMYSY 143
 :|:|:|:|:|

RESULT 5
 Q8EWD4 PRELIMINARY; PRT; 1123 AA.
 ID Q8EWD4
 AC Q8EWD4;
 DT 01-MAR-2003 (TReMBLrel. 23, Created)
 DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE MYPE 2560 paralog, 57%.
 GN MYPE2710.
 OS Mycoplasma penetrans.
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmaataceae; Mycoplasma.
 OX NCBI_TaxID=28227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HF-2;
 RX MEDLINE=22354719; PubMed=12466555;
 RA Sasaki Y., Ishikawa J., Yamashita A., Oshima K., Kenri T., Furuya K.,
 RA Yoshino C., Horino A., Shiba T., Sasaki T., Hattori M.;
 RT "The complete genomic sequence of Mycoplasma penetrans, an
 RT intracellular bacterial pathogen in humans.";
 RL Nucleic Acids Res. 30:5293-5300(2002).
 DR EMBL; AP004171; BAC44062.1; -;
 KW Complete proteome.
 SQ SEQUENCE 1123 AA; 123636 MW; A4D70730E3DB4AC CRC64;

Query Match 71.2%; Score 37; DB 16; Length 1123;
 Best Local Similarity 70.0%; Pred. No. 70;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 2 EVVPXGMSYS 11
 Db 658 EYVPMGLSYS 667
 :|:|:|:|:|

```

RESULT 6
Q8R126 PRELIMINARY; PRT; 471 AA.
ID Q8R126 AC Q8R126;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical 54.5 kDa protein (Fragment).
GN GLTSCR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
RS SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Strauberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC025810; AAH25810.1; -.
DR MGD; MGI:215441; Gltsr2.
KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 471 AA; 54506 MW; EODA685C374A9760 CRC64;

Query Match 69.2%; Score 36; DB 11; Length 471;
Best Local Similarity 60.0%; Pred.No. 44;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 EVVFXGMSYS 11
Db 226 EVIPAGASYN 235
||| |||

RESULT 7
Q8VD18 PRELIMINARY; PRT; 484 AA.
ID Q8VD18 AC Q8VD18;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to glioma tumor suppressor candidate region gene 2.
GN GLTSCR2 OR AW53641.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
RS SEQUENCE FROM N.A.
RC TISSUE=Salivary gland;
RA Strauberg R.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC017637; AAH17637.1; -.
DR MGD; MGI:215441; Gltsr2.
SQ SEQUENCE 484 AA; 55935 MW; BBB45F3B4BE02A36 CRC64;

Query Match 69.2%; Score 36; DB 11; Length 484;
Best Local Similarity 60.0%; Pred.No. 45;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 EVVFXGMSYS 11
Db 239 EVIPAGASYN 248
||| |||

RESULT 8
Q8BTX4 PRELIMINARY; PRT; 484 AA.
ID Q8BTX4 AC Q8BTX4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

```

OX NCBI_TaxID=4081;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VF36; TISSUE=Pistil;
 RX MEDLINE=95375233; PubMed=7647301;
 RA Milligan S.B., Gasser C.S.;
 RT "Nature and regulation of pistil-expressed genes in tomato.";
 RL Plant Mol. Biol. 28:691-711(1995).
 DR EMBL; U20592; AAA80497.1; -.
 DR InterPro; IPR002160; Kunitz_legume.
 DR Pfam; PF00197; Kunitz_legume; 1.
 DR PRINTS; PR00291; KUNITZINHESTR.
 DR ProDom; PD000891; Kunitz_legume; 1.
 DR SMART; SM00452; ST1; 1.
 DR PROSITE; PS00283; SOYBEAN KUNITZ; 1.
 KW Hypothetical protein; Signal.
 FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 225 UNKNOWN.
 SQ SEQUENCE 225 AA; 25188 MW; 1074C261D20CFDAD CRC64;

 Query Match 67.3%; Score 35; DB 10; Length 225;
 Best Local Similarity 54.5%; Pred. No. 32;
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

 Qy 1 BEVVPXGMSYS 11
 Db 32 DEVVPGKTYA 42

 RESULT 11
 Q9XVK4
 ID Q9XVK4 PRELIMINARY; PRT; 425 AA.
 AC Q9XVK4;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE R10D12.10 protein.
 GN R10D12.10
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoides;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Percy C.M.;
 RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 investigating biology.";
 RL Science 282:2012-2018(1998).
 DR EMBL; Z81109; CAB03241.1; -.
 DR WormPep; R10D12.10; CE12690.
 DR InterPro; IPR000719; Prot_kinase.
 DR Pfam; PF00069; pkinase; 1.
 DR ProDom; PD000001; Prot_kinase; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 KW ATP-binding; Transferase.
 SQ SEQUENCE 425 AA; 49410 MW; 5D96E29B08C8E9D6 CRC64;

Query Match 67.3%; Score 35; DB 5; Length 425;
 Best Local Similarity 50.0%; Pred. No. 64;
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

 Qy 1 BEVVPXGMSY 10
 Db 335 EQIVPGGLQY 344

 RESULT 12
 Q9C1D7

ID Q9C1D7 PRELIMINARY; PRT; 495 AA.
 AC Q9C1D7;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE CDNA FLJ31891 FIS.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CS7BL/6J; TISSUE=Head;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 DR EMBL; AK028258; BAC25846.1; -.
 SQ SEQUENCE 495 AA; 55358 MW; B6A22A093C114752 CRC64;

 Query Match 67.3%; Score 35; DB 11; Length 495;
 Best Local Similarity 63.6%; Pred. No. 75;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

 Qy 1 BEVVPXGMSYS 11
 Db 260 BEIVPMGISPS 270

 RESULT 13
 Q43733
 ID Q43733 PRELIMINARY; PRT; 556 AA.
 AC Q43733;
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
 DT 01-JUN-1998 (TrEMBLrel. 06, Last annotation update)
 DE DNA binding protein (Fragment).
 GN DX451B15.2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Tubby B.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; Z98050; CAB10847.1; -.
 FT NON TER 1
 SQ SEQUENCE 556 AA; 59059 MW; ECB00E4033FB2528 CRC64;

 Query Match 67.3%; Score 35; DB 4; Length 556;
 Best Local Similarity 66.7%; Pred. No. 86;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

 Qy 3 VVPXGMSYS 11
 Db 244 VVPAGLTYS 252

 RESULT 14
 Q9BH83
 ID Q9BH83 PRELIMINARY; PRT; 583 AA.
 AC Q9BH83;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Choline transporter.
 GN SCTL.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;

RN [1]
RP SEQUENCE FROM N.A.
RA Ben Mamoun C., Gluzman I.Y., Goldberg D.E.;
RT "Plasmodium falciparum choline transporter (PfSCT1) gene."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY007374; AAK14818.1; -
DR EMBL; AY007373; AAK14817.1; -
DR InterPro; IPR002123; Acyltransferase.
DR Pfam; PF01553; Acyltransferase; 1.
DR SMART; SM00563; PlsC; 1.
SQ SEQUENCE 583 AA; 66918 MW; 4DF83D7530527474 CRC64;

Query Match 67.3%; Score 35; DB 5; Length 583;
Best Local Similarity 55.6%; Pred. No. 90;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 3 VVPXGMSYS 11

DB 227 IIPVGLSYS 235

RESULT 15

Q9BHA5 PRELIMINARY; PRT; 583 AA.
AC Q9BHA5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Choline transporter.
GN SCT1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RA Ben Mamoun C., Gluzman I.Y., Goldberg D.E.;
RT "Plasmodium falciparum choline transporter (PfSCT1) gene."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY007372; AAK14816.1; -
DR EMBL; AY007375; AAK17947.1; -
DR InterPro; IPR002123; Acyltransferase.
DR Pfam; PF01553; Acyltransferase; 1.
DR SMART; SM00563; PlsC; 1.
SQ SEQUENCE 583 AA; 66917 MW; 2B2BFAE3E395E049 CRC64;

Query Match 67.3%; Score 35; DB 5; Length 583;
Best Local Similarity 55.6%; Pred. No. 90;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 3 VVPXGMSYS 11

DB 227 IIPVGLSYS 235

Search completed: December 22, 2003, 17:51:29
Job time : 25.2667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:41:00 ; Search time 32.4667 Seconds
(without alignments)
53.778 Million cell updates/sec

Title: US-09-909-164-11

Perfect score: 56

Sequence: 1 EVVVP²GHYS(11)

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 159726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq 19Jun03.*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
- 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|-------|-------|--------|----|----------|-------------------|
| 1 | 54 | 96.4 | 11 | 23 | ABB80523 | Hepatitis C virus |
| 2 | 54 | 96.4 | 11 | 23 | ABB80527 | Hepatitis C virus |
| 3 | 54 | 96.4 | 11 | 23 | ABB80558 | Hepatitis C virus |
| 4 | 54 | 96.4 | 11 | 23 | ABB80560 | Hepatitis C virus |
| 5 | 49 | 87.5 | 11 | 23 | ABB80537 | Hepatitis C virus |
| 6 | 49 | 87.5 | 11 | 23 | ABB80541 | Hepatitis C virus |
| 7 | 48 | 85.7 | 11 | 23 | ABB80546 | Hepatitis C virus |
| 8 | 48 | 85.7 | 11 | 23 | ABB80550 | Hepatitis C virus |
| 9 | 48 | 85.7 | 11 | 23 | ABB80554 | Hepatitis C virus |

| | | | | | | |
|----|----|------|----|----|----------|-------------------|
| 10 | 48 | 85.7 | 11 | 23 | ABB80555 | Hepatitis C virus |
| 11 | 46 | 82.1 | 11 | 23 | ABB80531 | Hepatitis C virus |
| 12 | 46 | 82.1 | 11 | 23 | ABB80532 | Hepatitis C virus |
| 13 | 45 | 80.4 | 11 | 23 | ABB80521 | Hepatitis C virus |
| 14 | 45 | 80.4 | 11 | 23 | ABB80522 | Hepatitis C virus |
| 15 | 45 | 80.4 | 11 | 23 | ABB80524 | Hepatitis C virus |
| 16 | 45 | 80.4 | 11 | 23 | ABB80525 | Hepatitis C virus |
| 17 | 45 | 80.4 | 11 | 23 | ABB80526 | Hepatitis C virus |
| 18 | 45 | 80.4 | 11 | 23 | ABB80528 | Hepatitis C virus |
| 19 | 45 | 80.4 | 11 | 23 | ABB80529 | Hepatitis C virus |
| 20 | 45 | 80.4 | 11 | 23 | ABB80529 | Hepatitis C virus |
| 21 | 45 | 80.4 | 11 | 23 | ABB80529 | Hepatitis C virus |
| 22 | 45 | 80.4 | 11 | 23 | ABB80561 | Hepatitis C virus |
| 23 | 45 | 80.4 | 11 | 23 | ABB80562 | Hepatitis C virus |
| 24 | 45 | 80.4 | 11 | 23 | ABB80563 | Hepatitis C virus |
| 25 | 45 | 80.4 | 11 | 23 | ABB80564 | Hepatitis C virus |
| 26 | 45 | 80.4 | 11 | 23 | ABB80565 | Hepatitis C virus |
| 27 | 45 | 80.4 | 11 | 23 | ABB80566 | Hepatitis C virus |
| 28 | 45 | 80.4 | 11 | 23 | ABB80567 | Hepatitis C virus |
| 29 | 40 | 71.4 | 11 | 23 | ABB80535 | Hepatitis C virus |
| 30 | 40 | 71.4 | 11 | 23 | ABB80536 | Hepatitis C virus |
| 31 | 40 | 71.4 | 11 | 23 | ABB80538 | Hepatitis C virus |
| 32 | 40 | 71.4 | 11 | 23 | ABB80539 | Hepatitis C virus |
| 33 | 40 | 71.4 | 11 | 23 | ABB80540 | Hepatitis C virus |
| 34 | 40 | 71.4 | 11 | 23 | ABB80542 | Hepatitis C virus |
| 35 | 40 | 71.4 | 11 | 23 | ABB80543 | Hepatitis C virus |
| 36 | 39 | 69.6 | 11 | 23 | ABB80544 | Hepatitis C virus |
| 37 | 39 | 69.6 | 11 | 23 | ABB80545 | Hepatitis C virus |
| 38 | 39 | 69.6 | 11 | 23 | ABB80547 | Hepatitis C virus |
| 39 | 39 | 69.6 | 11 | 23 | ABB80548 | Hepatitis C virus |
| 40 | 39 | 69.6 | 11 | 23 | ABB80549 | Hepatitis C virus |
| 41 | 39 | 69.6 | 11 | 23 | ABB80551 | Hepatitis C virus |
| 42 | 39 | 69.6 | 11 | 23 | ABB80552 | Hepatitis C virus |
| 43 | 39 | 69.6 | 11 | 23 | ABB80553 | Hepatitis C virus |
| 44 | 39 | 69.6 | 11 | 23 | ABB80556 | Hepatitis C virus |
| 45 | 39 | 69.6 | 11 | 23 | ABB80557 | Hepatitis C virus |

ALIGNMENTS

| | | | | | | |
|----------|------------------------------|--|--|----|-----|--|
| RESULT 1 | | | | | | |
| ABB80523 | | | | | | |
| ID | ABB80523 | standard; | peptide; | 11 | AA. | |
| XX | ABB80523; | | | | | |
| AC | ABB80523; | | | | | |
| XX | 08-OCT-2002 | (first entry) | | | | |
| DT | Hepatitis C virus NS3/NS4a | serine protease inhibitor peptide #3. | | | | |
| DE | Hepatitis C virus; HCV; | serine protease; inhibitor; alpha-ketoamide; | | | | |
| XX | Hepatitis C virus; HCV; | serine protease; inhibitor; alpha-ketoamide; | | | | |
| KW | virucide. | | | | | |
| XX | Synthetic. | | | | | |
| OS | Synthetic. | | | | | |
| XX | Key | Location/Qualifiers | | | | |
| FT | Modified-site | 1 | /note= "N-terminal acetyl" | | | |
| FT | Modified-site | 6 | /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7" | | | |
| FT | Misc-difference | 9 | /note= "D-form residue" | | | |
| FT | Modified-site | 11 | /note= "C-terminal amide" | | | |
| XX | WO200208251-A2. | | | | | |
| XX | 31-JAN-2002. | | | | | |
| PD | 19-JUL-2001; 2001WO-US23169. | | | | | |
| PF | | | | | | |
| XX | | | | | | |

PR 21-JUL-2000; 2000US-220101P.
 PA (CORV-) CORVAS INT INC.
 XX Lim-wilby M, Levy OE, Brunck TK;
 XX WPI; 2002-361643/39.
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C
 PT virus protease -
 XX Claim 17; Page 64; 69pp; English.
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.
 XX Sequence 11 AA;
 SQ Query Match 96.4%; Score 54; DB 23; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.00053;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 EEVVPXGMHYS 11
 |||||
 DB 1 EEVVPXGMHYS 11
 |||||
 RESULT 2
 ABB80527
 ID ABB80527 standard; peptide; 11 AA.
 XX ABB80527;
 XX 08-OCT-2002 (first entry)
 XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #7.
 XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.
 XX Synthetic.
 OS Key Location/Qualifiers
 FH Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Misc-difference 8 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"
 FT WO200208251-A2.
 XX 31-JAN-2002.
 XX 19-JUL-2001; 2001WO-US23169.
 XX 21-JUL-2000; 2000US-220101P.
 XX (CORV-) CORVAS INT INC.
 XX Lim-wilby M, Levy OE, Brunck TK;
 XX WPI; 2002-361643/39.
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C
 PT virus protease -
 XX Claim 17; Page 65; 69pp; English.
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.
 XX Sequence 11 AA;
 SQ Query Match 96.4%; Score 54; DB 23; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.00053;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 EEVVPXGMHYS 11
 |||||
 DB 1 EEVVPXGMHYS 11
 |||||

PT Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C
 PT virus protease -
 XX Claim 17; Page 64; 69pp; English.
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.
 XX Sequence 11 AA;
 SQ Query Match 96.4%; Score 54; DB 23; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.00053;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 EEVVPXGMHYS 11
 |||||
 DB 1 EEVVPXGMHYS 11
 |||||
 RESULT 3
 ABB80558
 ID ABB80558 standard; peptide; 11 AA.
 XX ABB80558;
 XX 08-OCT-2002 (first entry)
 XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #38.
 XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.
 XX Synthetic.
 OS Key Location/Qualifiers
 FH Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Modified-site 8 /note= "Oxymethionine"
 FT Modified-site 11 /note= "C-terminal amide"
 FT WO200208251-A2.
 XX 31-JAN-2002.
 XX 19-JUL-2001; 2001WO-US23169.
 XX 21-JUL-2000; 2000US-220101P.
 XX (CORV-) CORVAS INT INC.
 XX Lim-wilby M, Levy OE, Brunck TK;
 XX WPI; 2002-361643/39.
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C
 PT virus protease -
 XX Claim 17; Page 65; 69pp; English.
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the

CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.

XX SQ Sequence 11 AA;
 Query Match 96.4%; Score 54; DB 23; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.00053; Mismatches 0; Indels 0; Gaps 0;
 Matches 11; Conservative 0;

OY 1 BEVVPXGMHYS 11
 |||||
 DB 1 BEVVPXGMHYS 11

RESULT 4
 ABB80560
 ID ABB80560 standard; peptide; 11 AA.
 XX AC ABB80560;
 XX DT 08-OCT-2002 (first entry)
 DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #40.
 XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 XX virucide.
 XX OS Synthetic.

XX FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Misc-difference 8 /note= "D-form residue"
 FT Modified-site 8 /note= "Oxymethionine"
 FT Misc-difference 9 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"

XX PN WO200208251-A2.
 XX PD 31-JAN-2002.
 XX PF 19-JUL-2001; 2001WO-US23169.
 XX PR 21-JUL-2000; 2000US-220101P.
 XX PA (CORV-) CORVAS INT INC.

XX PI Lin-wilby M, Levy OE, Brunck TK;
 XX DR WPI; 2002-361643/39.
 XX PT Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C
 PT virus protease -
 XX PS Claim 17; Page 65; 69pp; English.

CC The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient

CC is useful for treating disorders associated with hepatitis C virus.
 XX SQ Sequence 11 AA;

Query Match 96.4%; Score 54; DB 23; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.00053; Mismatches 0; Indels 0; Gaps 0;
 Matches 11; Conservative 0;

OY 1 BEVVPXGMHYS 11
 |||||
 DB 1 BEVVPXGMHYS 11

RESULT 5
 ABB80537
 ID ABB80537 standard; peptide; 11 AA.
 XX AC ABB80537;
 XX DT 08-OCT-2002 (first entry)
 DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #17.
 XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 XX virucide.
 XX OS Synthetic.

XX FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Misc-difference 9 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"
 FT WO200208251-A2.
 XX PD 31-JAN-2002.

XX PF 19-JUL-2001; 2001WO-US23169.
 XX PR 21-JUL-2000; 2000US-220101P.
 XX PA (CORV-) CORVAS INT INC.

XX PI Lin-wilby M, Levy OE, Brunck TK;
 XX DR WPI; 2002-361643/39.

XX PT Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C
 PT virus protease -

XX PS Claim 17; Page 64; 69pp; English.

CC The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.

XX SQ Sequence 11 AA;

Query Match 87.5%; Score 49; DB 23; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.0051; Mismatches 10; Conservative 0; Indels 1; Gaps 0;

OY 1 BEVVPXGMHYS 11
 |||||
 DB 1 BEVVPXGMHYS 11

RESULT 6

ABB80541
 ID ABB80541 standard; peptide; 11 AA.

AC ABB80541;
 DT 08-OCT-2002 (first entry)
 DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #21.
 KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 virucide.
 XX Synthetic.
 XX Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 residue 7"
 FT Misc-difference 8 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"
 FT WO200208251-A2.
 PN 31-JAN-2002.
 XX 19-JUL-2001; 2001WO-US23169.
 XX 21-JUL-2000; 2000US-220101P.
 XX (CORV-) CORVAS INT INC.
 PI Lim-wilby M, Levy OE, Brunck TK;
 XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitor
 PT activity useful for treating disorders associated with hepatitis C
 PT virus protease
 XX Claim 17; Page 65; 69pp; English.
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.

PS Claim 17; Page 65; 69pp; English.

XX Query Match 87.5%; Score 49; DB 23; Length 11;
 CC Best Local Similarity 90.9%; Pred. No. 0.0051;
 CC Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

XX Sequence 11 AA;

OY 1 BEVVPXGMHYS 11
 |||||
 DB 1 BEVVPXGMHYS 11

RESULT 7

ABB80546
 ID ABB80546 standard; peptide; 11 AA.

XX ABB80546;
 AC 08-OCT-2002 (first entry)
 DT Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #26.
 DE Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 virucide.
 XX Synthetic.
 XX Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 residue 7"
 FT Modified-site 11 /note= "C-terminal amide"
 FT WO200208251-A2.
 PN 31-JAN-2002.
 XX 19-JUL-2001; 2001WO-US23169.
 XX 21-JUL-2000; 2000US-220101P.
 XX (CORV-) CORVAS INT INC.
 PI Lim-wilby M, Levy OE, Brunck TK;
 XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitor
 PT activity useful for treating disorders associated with hepatitis C
 PT virus protease
 XX Claim 17; Page 65; 69pp; English.

XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.

XX Sequence 11 AA;

Query Match 85.7%; Score 48; DB 23; Length 11;
 CC Best Local Similarity 90.9%; Pred. No. 0.008;
 CC Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 BEVVPXGMHYS 11
 |||||
 DB 1 BEVVPXGMHYS 11

RESULT 8

ABB80550
 ID ABB80550 standard; peptide; 11 AA.

AC ABB80550;
 DT 08-OCT-2002 (first entry)
 DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #30.
 KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 virucide.

RESULT 9

ABB80546
 ID ABB80546 standard; peptide; 11 AA.

OS Synthetic.
 XX Key Location/Qualifiers
 FH Modified-site 1
 FT /note= "N-terminal acetyl"
 FT Modified-site 6
 FT /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Misc-difference 9
 FT /note= "D-form residue"
 FT Modified-site 11
 FT /note= "C-terminal amide"
 XX WO200208251-A2.
 PN 31-JAN-2002.
 XX 19-JUL-2001; 2001WO-US23169.
 XX 21-JUL-2000; 2000US-220101P.
 PR (CORV-) CORVAS INT INC.
 XX Lim-wilby M, Levy OE, Brunck TK;
 PI WPI; 2002-361643/39.
 DR Novel peptide compound having hepatitis C virus protease inhibitory
 XX activity useful for treating disorders associated with hepatitis C
 XX virus protease -
 XX Claim 17; Page 65; 69pp; English.
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.
 XX SQ Sequence 11 AA;
 Query Match 85.7%; Score 48; DB 23; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.008;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 BEVVPXGMHYS 11
 Db ||||| |||||
 1 BEVVPXGSHYS 11
 RESULT 9
 ABB80554
 ID ABB80554 standard; peptide; 11 AA.
 XX AC ABB80554;
 XX 08-OCT-2002 (first entry)
 DT Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #34.
 DE Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 XX virucide.
 XX Synthetic.
 OS Key Location/Qualifiers
 FH Modified-site 1
 FT /note= "N-terminal acetyl"
 FT Modified-site 6
 FT /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Misc-difference 9
 FT /note= "D-form residue"
 FT Modified-site 11
 FT /note= "C-terminal amide"
 XX WO200208251-A2.
 PN 31-JAN-2002.
 XX 19-JUL-2001; 2001WO-US23169.
 XX 21-JUL-2000; 2000US-220101P.
 PR (CORV-) CORVAS INT INC.
 XX Lim-wilby M, Levy OE, Brunck TK;
 PI WPI; 2002-361643/39.
 DR Novel peptide compound having hepatitis C virus protease inhibitory
 XX activity useful for treating disorders associated with hepatitis C
 XX virus protease -
 XX Claim 17; Page 65; 69pp; English.
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.
 XX SQ Sequence 11 AA;
 Query Match 85.7%; Score 48; DB 23; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.008;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 BEVVPXGMHYS 11
 Db ||||| |||||
 1 BEVVPXGSHYS 11
 RESULT 9
 ABB80554
 ID ABB80554 standard; peptide; 11 AA.
 XX AC ABB80554;
 XX 08-OCT-2002 (first entry)
 DT Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #34.
 DE Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 XX virucide.
 XX Synthetic.
 OS Key Location/Qualifiers
 FH Modified-site 1
 FT /note= "N-terminal acetyl"
 FT Modified-site 6
 FT /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Misc-difference 9
 FT /note= "D-form residue"
 FT Modified-site 11
 FT /note= "C-terminal amide"
 XX WO200208251-A2.
 PN 31-JAN-2002.
 XX 19-JUL-2001; 2001WO-US23169.
 XX 21-JUL-2000; 2000US-220101P.
 PR (CORV-) CORVAS INT INC.
 XX Lim-wilby M, Levy OE, Brunck TK;
 PI WPI; 2002-361643/39.
 DR Novel peptide compound having hepatitis C virus protease inhibitory
 XX activity useful for treating disorders associated with hepatitis C
 XX virus protease -
 XX Claim 17; Page 65; 69pp; English.
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.
 XX SQ Sequence 11 AA;
 Query Match 85.7%; Score 48; DB 23; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.008;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 BEVVPXGMHYS 11
 Db ||||| |||||
 1 BEVVPXGSHYS 11
 RESULT 10
 ABB80555
 ID ABB80555 standard; peptide; 11 AA.
 XX AC ABB80555;
 XX 08-OCT-2002 (first entry)
 DT Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #35.
 DE Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 XX virucide.
 XX Synthetic.
 OS Key Location/Qualifiers
 FH Modified-site 1
 FT /note= "N-terminal acetyl"
 FT Modified-site 6
 FT /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Misc-difference 8
 FT /note= "D-form residue"
 FT Misc-difference 9
 FT /note= "D-form residue"
 FT Modified-site 11
 FT /note= "C-terminal amide"
 XX WO200208251-A2.
 PN 31-JAN-2002.
 XX 19-JUL-2001; 2001WO-US23169.
 XX 21-JUL-2000; 2000US-220101P.
 PR (CORV-) CORVAS INT INC.
 XX Lim-wilby M, Levy OE, Brunck TK;
 PI WPI; 2002-361643/39.
 DR Novel peptide compound having hepatitis C virus protease inhibitory
 XX activity useful for treating disorders associated with hepatitis C
 XX virus protease -
 XX Claim 17; Page 65; 69pp; English.
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.
 XX SQ Sequence 11 AA;
 Query Match 85.7%; Score 48; DB 23; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.008;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 BEVVPXGMHYS 11
 Db ||||| |||||
 1 BEVVPXGSHYS 11

FT Misc-difference 8
 FT /note= "D-form residue"
 FT Modified-site 11
 FT /note= "C-terminal amide"
 XX WO200208251-A2.
 PN 31-JAN-2002.
 XX 19-JUL-2001; 2001WO-US23169.
 XX 21-JUL-2000; 2000US-220101P.
 PR (CORV-) CORVAS INT INC.
 XX Lim-wilby M, Levy OE, Brunck TK;
 PI WPI; 2002-361643/39.
 DR Novel peptide compound having hepatitis C virus protease inhibitory
 XX activity useful for treating disorders associated with hepatitis C
 XX virus protease -
 XX Claim 17; Page 65; 69pp; English.
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.
 XX SQ Sequence 11 AA;
 Query Match 85.7%; Score 48; DB 23; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.008;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 BEVVPXGMHYS 11
 Db ||||| |||||
 1 BEVVPXGSHYS 11
 RESULT 10
 ABB80555
 ID ABB80555 standard; peptide; 11 AA.
 XX AC ABB80555;
 XX 08-OCT-2002 (first entry)
 DT Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #35.
 DE Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 XX virucide.
 XX Synthetic.
 OS Key Location/Qualifiers
 FH Modified-site 1
 FT /note= "N-terminal acetyl"
 FT Modified-site 6
 FT /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Misc-difference 8
 FT /note= "D-form residue"
 FT Misc-difference 9
 FT /note= "D-form residue"
 FT Modified-site 11
 FT /note= "C-terminal amide"
 XX WO200208251-A2.
 PN 31-JAN-2002.
 XX 19-JUL-2001; 2001WO-US23169.
 XX 21-JUL-2000; 2000US-220101P.
 PR (CORV-) CORVAS INT INC.
 XX Lim-wilby M, Levy OE, Brunck TK;
 PI WPI; 2002-361643/39.
 DR Novel peptide compound having hepatitis C virus protease inhibitory
 XX activity useful for treating disorders associated with hepatitis C
 XX virus protease -
 XX Claim 17; Page 65; 69pp; English.
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.
 XX SQ Sequence 11 AA;
 Query Match 85.7%; Score 48; DB 23; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.008;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 BEVVPXGMHYS 11
 Db ||||| |||||
 1 BEVVPXGSHYS 11

XX PD 31-JAN-2002.
XX PF 19-JUL-2001; 2001WO-US23169.
XX PR 21-JUL-2000; 2000US-220101P.
XX PA (CORV-) CORVAS INT INC.
XX PI Lim-wilby M, Levy OE, Brunck TK;
XX DR WPI; 2002-361643/39.
XX PT Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX PS Claim 17; Page 65; 69pp; English.
XX CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX SQ Sequence 11 AA;
XX Query Match 85.7%; Score 48; DB 23; Length 11;
XX Best Local Similarity 90.9%; Pred. No. 0.008;
XX Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX QY 1 EEVVPXGMHYS 11
XX DB ||||| |||||
XX 1 EEVVPXGMHYS 11
XX RESULT 11
XX ABB80531
XX ID ABB80531 standard; peptide; 11 AA.
XX AC ABB80531;
XX DT 08-OCT-2002 (first entry)
XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #11.
XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX OS Synthetic.
XX FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Modified-site 11 /note= "C-terminal amide"
XX WO200208251-A2.
XX 31-JAN-2002.
XX PD 19-JUL-2001; 2001WO-US23169.
XX PF 21-JUL-2000; 2000US-220101P.
XX PR (CORV-) CORVAS INT INC.
XX PI Lim-wilby M, Levy OE, Brunck TK;

XX WPI; 2002-361643/39.
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX PS Claim 17; Page 64; 69pp; English.
XX CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX SQ Sequence 11 AA;
XX Query Match 82.1%; Score 46; DB 23; Length 11;
XX Best Local Similarity 90.9%; Pred. No. 0.02;
XX Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX QY 1 EEVVPXGMHYS 11
XX DB ||||| |||||
XX 1 EEVVPXGMHYS 11
XX RESULT 12
XX ABB80532
XX ID ABB80532 standard; peptide; 11 AA.
XX AC ABB80532;
XX DT 08-OCT-2002 (first entry)
XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #12.
XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX OS Synthetic.
XX FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX WO200208251-A2.
XX 31-JAN-2002.
XX PD 19-JUL-2001; 2001WO-US23169.
XX PF 21-JUL-2000; 2000US-220101P.
XX PR (CORV-) CORVAS INT INC.
XX PA Lim-wilby M, Levy OE, Brunck TK;
XX PI WPI; 2002-361643/39.
XX DR Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX PS Claim 17; Page 64; 69pp; English.

XX CC The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.

XX SQ Sequence 11 AA;

Query Match 82.1%; Score 46; DB 23; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.02; Mismatches 0; Indels 0; Gaps 0;
 Matches 10; Conservative 0;

Qy 1 BEVVPXGMHYS 11
 |||||
 Db 1 BEVVPXGGHYS 11

RESULT 13

ABB80521
 ID ABB80521 standard; peptide; 11 AA.

XX AC ABB80521;

XX DT 08-OCT-2002 (first entry)

XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #1.

XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 virucide.

XX OS Synthetic.

XX FH Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 residue 7"

FT Modified-site 11 /note= "C-terminal amide"

XX WO200208251-A2.

XX 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US233169.

XX 21-JUL-2000; 2000US-220101P.

XX (CORV-) CORVAS INT INC.

XX Lim-wilby M, Levy OE, Brunck TK;

XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory
 activity useful for treating disorders associated with hepatitis C
 virus protease -

XX Claim 17; Page 64; 69pp; English.

XX CC The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.

XX SQ Sequence 11 AA;

Query Match 82.1%; Score 46; DB 23; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.02; Mismatches 0; Indels 0; Gaps 0;
 Matches 10; Conservative 0;

Qy 1 BEVVPXGMHYS 11
 |||||
 Db 1 BEVVPXGGHYS 11

RESULT 14

ABB80522
 ID ABB80522 standard; peptide; 11 AA.

XX AC ABB80522;

XX DT 08-OCT-2002 (first entry)

XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #2.

XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 virucide.

XX OS Synthetic.

XX FH Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 residue 7"

FT Misc-difference 9 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

XX WO200208251-A2.

XX 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US233169.

XX 21-JUL-2000; 2000US-220101P.

XX (CORV-) CORVAS INT INC.

XX Lim-wilby M, Levy OE, Brunck TK;

XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory
 activity useful for treating disorders associated with hepatitis C
 virus protease -

XX Claim 17; Page 64; 69pp; English.

XX CC The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.

XX SQ Sequence 11 AA;

Query Match 80.4%; Score 45; DB 23; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.031; Mismatches 0; Indels 0; Gaps 0;
 Matches 10; Conservative 0;

Qy 1 BEVVPXGMHYS 11
 |||||
 Db 1 BEVVPXGMSYS 11


```

RESULT 15
ABB0524
ID ABB0524 standard; peptide; 11 AA.
XX
AC ABB0524;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #4.
XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
virucide.
XX
OS Synthetic.
XX
FH
FT Key Location/Qualifiers
FT Modified-site 1
FT /note= "N-terminal acetyl"
FT Modified-site 6
FT /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 9
FT /note= "D-form residue"
FT Modified-site 11
FT /note= "C-terminal amide"
XX
PN WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US23169.
XX
XX 21-JUL-2000; 2000US-220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
XX Claim 17; Page 64; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;
Query Match 80.4%; Score 45; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.031;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGMHYS 11
DB 1 EEVVPXGMHYS 11

```

Search completed: December 22, 2003, 17:41:01
Job time : 32.4667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:37:03 ; Search time 10.1333 Seconds
(without alignments)
45.930 Million cell updates/sec

Title: US-09-909-164-11
Perfect score: 56
Sequence: 1 BEVVPXGMHYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2.6/prodata/1/iaa/5A_COMB.pep:*
2: /cgn2.6/prodata/1/iaa/5B_COMB.pep:*
3: /cgn2.6/prodata/1/iaa/6A_COMB.pep:*
4: /cgn2.6/prodata/1/iaa/6B_COMB.pep:*
5: /cgn2.6/prodata/1/iaa/PCUS_COMB.pep:*
6: /cgn2.6/prodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|----------------------|
| 1 | 39 | 69.6 | 1037 | 4 | US-09-134-001C-4794 |
| 2 | 37 | 66.1 | 856 | 4 | US-09-252-991A-21444 |
| 3 | 34 | 60.7 | 600 | 2 | US-08-821-119-19 |
| 4 | 34 | 60.7 | 600 | 2 | US-08-821-118-2 |
| 5 | 33 | 58.9 | 277 | 4 | US-09-252-991A-26615 |
| 6 | 33 | 58.9 | 385 | 4 | US-09-252-991A-27834 |
| 7 | 33 | 58.9 | 747 | 4 | US-09-724-864-36 |
| 8 | 32 | 57.1 | 70 | 4 | US-09-134-001C-3950 |
| 9 | 32 | 57.1 | 102 | 2 | US-08-580-988A-23 |
| 10 | 32 | 57.1 | 126 | 2 | US-08-879-995A-3 |
| 11 | 32 | 57.1 | 126 | 2 | US-09-215-096-3 |
| 12 | 32 | 57.1 | 152 | 2 | US-08-460-694-4 |
| 13 | 32 | 57.1 | 152 | 3 | US-08-460-744-4 |
| 14 | 32 | 57.1 | 152 | 3 | US-07-667-711B-4 |
| 15 | 32 | 57.1 | 173 | 1 | US-08-193-977-7 |
| 16 | 32 | 57.1 | 189 | 2 | US-08-464-517-21 |
| 17 | 32 | 57.1 | 189 | 2 | US-08-246-361A-21 |
| 18 | 32 | 57.1 | 189 | 3 | US-08-463-772-21 |
| 19 | 32 | 57.1 | 189 | 5 | PCT-US93-05000-21 |
| 20 | 32 | 57.1 | 236 | 2 | US-08-464-517-22 |
| 21 | 32 | 57.1 | 236 | 2 | US-08-246-361A-22 |
| 22 | 32 | 57.1 | 236 | 3 | US-08-463-772-22 |
| 23 | 32 | 57.1 | 236 | 5 | PCT-US93-05000-22 |
| 24 | 32 | 57.1 | 280 | 2 | US-08-464-517-6 |
| 25 | 32 | 57.1 | 280 | 3 | US-08-463-772-6 |
| 26 | 32 | 57.1 | 289 | 2 | US-08-246-361A-4 |
| 27 | 32 | 57.1 | 289 | 5 | PCT-US93-05000-4 |

28 57.1 291 5 PCT-US93-05000-6 Sequence 6, Appli
29 57.1 292 2 US-08-464-517-23 Sequence 23, Appli
30 57.1 292 2 US-08-246-361A-6 Sequence 6, Appli
31 57.1 292 2 US-08-246-361A-23 Sequence 23, Appli
32 57.1 292 3 US-08-463-772-23 Sequence 23, Appli
33 57.1 292 5 PCT-US93-05000-23 Sequence 23, Appli
34 57.1 295 1 US-07-947-120-8 Sequence 8, Appli
35 57.1 295 1 US-08-472-893A-8 Sequence 8, Appli
36 57.1 295 2 US-08-460-694-2 Sequence 19, Appli
37 57.1 295 2 US-08-464-517-19 Sequence 19, Appli
38 57.1 295 2 US-08-464-517-19 Sequence 20, Appli
39 57.1 295 2 US-08-246-361A-19 Sequence 20, Appli
40 57.1 295 3 US-08-246-361A-20 Sequence 20, Appli
41 57.1 295 3 US-08-463-772-19 Sequence 19, Appli
42 57.1 295 3 US-08-463-772-20 Sequence 19, Appli
43 57.1 295 3 US-08-460-744-2 Sequence 2, Appli
44 57.1 295 3 US-07-667-711B-2 Sequence 2, Appli
45 57.1 295 3 US-08-947-492-8 Sequence 8, Appli

ALIGNMENTS

RESULT 1

US-09-134-001C-4794
; Sequence 4794, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCI
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4794
; LENGTH: 1037
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4794

Query Match 69.6%; Score 39; DB 4; Length 1037;
Best Local Similarity 63.6%; Pred. No. 20;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Oy 1 BEVVPXGMHYS 11

Db 199 KEVWSNGLHYS 209

RESULT 2

US-09-252-991A-21444
; Sequence 21444, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21444
; LENGTH: 856
; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21444

Query Match 66.1%; Score 37; DB 4; Length 856;
Best Local Similarity 70.0%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMHY 10
DB 64 EAVVGGSEHY 73

RESULT 3

US-08-821-119-19
; Sequence 19, Application US/08821119
; Patent No. 5821104
; GENERAL INFORMATION:
; APPLICANT: Holm, Kaj Andre
; APPLICANT: Rasmussen, Grethe
; APPLICANT: Halkier, Torben
; APPLICANT: Lehmebeck, Jan
; TITLE OF INVENTION: Tripeptidyl Aminopeptidase
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5821104o No. 5821104disk of No. 5821104th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/821,119
; FILING DATE: 19-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4107.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 600 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-08-821-119-19

Query Match 60.7%; Score 34; DB 2; Length 600;
Best Local Similarity 75.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 VPXGMHY 11
DB 31 VPKGMHY 38

RESULT 4

US-08-821-118-2
; Sequence 2, Application US/08821118
; Patent No. 5989889
; GENERAL INFORMATION:
; APPLICANT: Rey, Michael
; APPLICANT: Golightly, Elizabeth

; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING POLYPEPTIDES
; TITLE OF INVENTION: HAVING TRIPEPTIDE AMINOPEPTIDASE
; TITLE OF INVENTION: ACTIVITY

; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5989889o No. 5989889disk of No. 5989889th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/821,118
; FILING DATE: 19-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4107.400-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 600 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-08-821-118-2

Query Match 60.7%; Score 34; DB 2; Length 600;
Best Local Similarity 75.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 VPXGMHY 11
DB 31 VPKGMHY 38

RESULT 5

US-09-252-991A-26615
; Sequence 26615, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26615
; LENGTH: 277
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26615

Query Match 58.9%; Score 33; DB 4; Length 277;
Best Local Similarity 63.6%; Pred. No. 70;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 BEVVPXGMHYS 11
 |||||
 Db 48 BEVPGGHTS 58

RESULT 6

US-09-252-991A-27834
 ; Sequence 27834, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US/09/252,991A
 ; PRIOR FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 27834
 ; LENGTH: 385
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-27834

Query Match 58.9%; Score 33; DB 4; Length 385;
 Best Local Similarity 44.4%; Pred. No. 1e+02; 2; Indels 0; Gaps 0;
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 RWPXGMHY 10
 |||||
 Db 201 BILPAALHY 209

RESULT 7

US-09-724-864-36
 ; Sequence 36, Application US/09724864
 ; Patent No. 6380362
 ; GENERAL INFORMATION:
 ; APPLICANT: Watson, James D
 ; APPLICANT: Murison, James G.
 ; TITLE OF INVENTION: Polynucleotides, polypeptides expressed
 ; TITLE OF INVENTION: by the polynucleotides and methods for their use.
 ; FILE REFERENCE: 11000.1050U1
 ; CURRENT APPLICATION NUMBER: US/09/724,864
 ; CURRENT FILING DATE: 2000-11-28
 ; PRIOR APPLICATION NUMBER: U.S. No. 6380362 60/171,678
 ; PRIOR FILING DATE: 1999-12-23
 ; NUMBER OF SEQ ID NOS: 72
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 36
 ; LENGTH: 747
 ; TYPE: PRT
 ; ORGANISM: Rat
 US-09-724-864-36

Query Match 58.9%; Score 33; DB 4; Length 747;
 Best Local Similarity 71.4%; Pred. No. 2.1e+02;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 5 PXGMHYS 11
 |||||
 Db 627 PGLHYS 633

RESULT 8

US-09-134-001C-3950
 ; Sequence 3950, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; CURRENT FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 3950
 ; LENGTH: 70
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-3950

Query Match 57.1%; Score 32; DB 4; Length 70;
 Best Local Similarity 62.5%; Pred. No. 24;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 VPXGMHYS 11
 |||||
 Db 36 MPXGMHYS 43

RESULT 9

US-08-580-988A-23
 ; Sequence 23, Application US/08580988A
 ; Patent No. 5856161
 ; GENERAL INFORMATION:
 ; APPLICANT: Aggarwal et al.
 ; TITLE OF INVENTION: Tumor Necrosis Factor
 ; TITLE OF INVENTION: Receptor-I-Associated Protein Kinase And Methods
 ; TITLE OF INVENTION: For Its Use
 ; NUMBER OF SEQUENCES: 27
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Dr. Benjamin A. Adler
 ; STREET: 8011 Candle Lane
 ; CITY: Houston
 ; STATE: Texas
 ; COUNTRY: USA
 ; ZIP: 77071
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 1.44 Mb floppy disk
 ; COMPUTER: Apple Macintosh
 ; OPERATING SYSTEM: Macintosh
 ; SOFTWARE: Microsoft Word for Macintosh
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/580,988A
 ; FILING DATE: January 3, 1996
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Benjamin Aaron Adler, Ph.D., J.D.
 ; REGISTRATION NUMBER: 35,423
 ; REFERENCE/DOCKET NUMBER: D5721CIP2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 713-777-2321
 ; TELEFAX: 713-777-6908
 ; INFORMATION FOR SEQ ID NO: 23:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 102 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS:
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; DESCRIPTION: no
 ; HYPOTHETICAL: no
 ; ANTI-SENSE: no
 ; FRAGMENT TYPE: internal
 ; ORIGINAL SOURCE:

US-08-580-988A-23

Query Match 57.1%; Score 32; DB 2; Length 102;
 Best Local Similarity 60.0%; Pred. No. 37;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVVPXGMH 10
 Db 24 BEVPLAWY 33

RESULT 10

US-08-879-995A-3
 ; Sequence 3, Application US/08879995A
 ; Patent No. 5985606
 ; GENERAL INFORMATION:
 ; APPLICANT: Hillman, Jennifer L.
 ; APPLICANT: Kaser, Matthew R.
 ; TITLE OF INVENTION: HUMAN PREPROTACHYKININ B
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.
 ; STREET: 3174 Porter Drive
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94304

COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/879,995A
 FILING DATE: Herewith
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Billings, Lucy J.
 REGISTRATION NUMBER: 36,749
 REFERENCE/DOCKET NUMBER: PF-0326 US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-855-0555
 TELEFAX: 415-845-4166
 TELEX:

INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 126 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 LIBRARY: GenBank
 CLONE: 163590
 US-08-879-995A-3

Query Match 57.1%; Score 32; DB 2; Length 126;
 Best Local Similarity 66.7%; Pred. No. 46;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 BEVVPXGMH 9
 Db 28 EQVPGGHH 36

RESULT 11

US-09-215-096-3
 ; Sequence 3, Application US/09215096
 ; Patent No. 6008194
 ; GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.
 APPLICANT: Lal, Preeti
 APPLICANT: Kaser, Matthew R.
 TITLE OF INVENTION: HUMAN PREPROTACHYKININ B
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Incyte Pharmaceuticals, Inc.
 STREET: 3174 Porter Drive
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94304
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/215,096
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/879,995
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Billings, Lucy J.
 REGISTRATION NUMBER: 36,749
 REFERENCE/DOCKET NUMBER: PF-0326 US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-855-0555
 TELEFAX: 415-845-4166
 TELEX:
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 126 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 LIBRARY: GenBank
 CLONE: 163590
 US-09-215-096-3

Query Match 57.1%; Score 32; DB 3; Length 126;
 Best Local Similarity 66.7%; Pred. No. 46;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 BEVVPXGMH 9
 Db 28 EQVPGGHH 36

RESULT 12

US-08-460-694-4
 ; Sequence 4, Application US/08460694
 ; Patent No. 5858655
 ; GENERAL INFORMATION:
 ; APPLICANT: Arnold, Andrew
 ; TITLE OF INVENTION: PRADI Cyclin and its cDNA
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
 ; STREET: 1100 New York Avenue, N.W., Suite 600
 ; CITY: Washington
 ; STATE: DC
 ; COUNTRY: USA
 ; ZIP: 20005

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/460,694
 FILING DATE: 02-JUN-1995
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: McConathy, Evelyn H.
 REGISTRATION NUMBER: 35,279
 REFERENCE/DOCKET NUMBER: 0609.4070002
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-371-2600
 TELEFAX: 202-371-2540
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 152 amino acids
 TYPE: amino acid
 STRANDEDNESS: not relevant
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-460-694-4

Query Match 57.1%; Score 32; DB 2; Length 152;
 Best Local Similarity 60.0%; Pred. No. 57;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 1 BEVVPXGMHY 10
 ||| ||| :|
 Db 20 BEVFPPLAMNY 29

RESULT 13
 US-08-460-744-4
 ; Sequence 4, Application US/08460744
 ; Patent No. 6107541
 ; GENERAL INFORMATION:
 ; APPLICANT: Arnold, Andrew
 ; TITLE OF INVENTION: PRAD1 Cyclin and its cDNA
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
 ; STREET: 1100 New York Avenue, N.W., Suite 600
 ; CITY: Washington
 ; STATE: DC
 ; COUNTRY: USA
 ; ZIP: 20005
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/460,744
 ; FILING DATE: 02-JUN-1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: McConathy, Evelyn H.
 ; REGISTRATION NUMBER: 35,279
 ; REFERENCE/DOCKET NUMBER: 0609.4070005
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-371-2600
 ; TELEFAX: 202-371-2540
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 152 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: not relevant
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 US-08-460-744-4

Query Match 57.1%; Score 32; DB 3; Length 152;
 Best Local Similarity 60.0%; Pred. No. 57;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 1 BEVVPXGMHY 10
 ||| ||| :|
 Db 20 BEVFPPLAMNY 29

RESULT 14
 US-07-667-711B-4
 ; Sequence 4, Application US/07667711B
 ; Patent No. 6110700
 ; GENERAL INFORMATION:
 ; APPLICANT: ARNOLD, ANDREW
 ; TITLE OF INVENTION: Prad1 Cyclin and its cDNA
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
 ; STREET: 1100 NEW YORK AVE., NW, SUITE 600
 ; CITY: WASHINGTON
 ; STATE: DC
 ; COUNTRY: USA
 ; ZIP: 20005
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/667,711B
 ; FILING DATE: 11-MAR-1991
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: MCPHAIL, DONALD R.
 ; REGISTRATION NUMBER: 35,811
 ; REFERENCE/DOCKET NUMBER: 0609.4070000
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 371-2600
 ; TELEFAX: (202) 371-2540
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 152 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: not relevant
 ; MOLECULE TYPE: peptide
 US-07-667-711B-4

Query Match 57.1%; Score 32; DB 3; Length 152;
 Best Local Similarity 60.0%; Pred. No. 57;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 1 BEVVPXGMHY 10
 ||| ||| :|
 Db 20 BEVFPPLAMNY 29

RESULT 15
 US-08-193-977-7
 ; Sequence 7, Application US/08193977
 ; Patent No. 5625031
 ; GENERAL INFORMATION:
 ; APPLICANT: WEBSTER, KEVIN R.
 ; APPLICANT: COLEMAN, KEVIN G.
 ; TITLE OF INVENTION: PEPTIDE INHIBITORS OF THE P33CDK2 AND
 ; TITLE OF INVENTION: P34CDC2 CELL CYCLE REGULATORY KINASES AND HUMAN
 ; TITLE OF INVENTION: PAPILLOMAVIRUS E7 ONCOPROTEIN
 ; NUMBER OF SEQUENCES: 34
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: REED & ROBINS
 ; STREET: 635 BRYANT STREET
 ; CITY: PALO ALTO
 ; STATE: CALIFORNIA
 ; COUNTRY: UNITED STATES OF AMERICA
 ; ZIP: 94301
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk

Query Match 57.1%; Score 32; DB 3; Length 152;
 Best Local Similarity 60.0%; Pred. No. 57;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 1 BEVVPXGMHY 10

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/193,977
FILING DATE: 08-FEB-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: ROBINS, ROBERTA L.
REGISTRATION NUMBER: 33,208
REFERENCE/DOCKET NUMBER: 5998-0016
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 617-8999
TELEFAX: (415) 327-3231
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 173 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-193-977-7

Query Match 57.1%; Score 32; DB 1; Length 173;
Best Local Similarity 60.0%; Pred. No. 65;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMHY 10
||| | | |
DB 55 BEVFPANNY 64

Search completed: December 22, 2003, 16:43:43
Job time : 10.1333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:40:14 ; Search time 20.8667 Seconds
(without alignments)
98.451 Million cell updates/sec

Title: US-09-909-164-11
Perfect score: 56
Sequence: 1 EEVVPXGMHYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 696363 seqs, 186758610 residues

Total number of hits satisfying chosen parameters: 696363

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|---------------------------------------|
| 1 | 33 | 58.9 | 273 | 12 | US-10-094-749-2319 Sequence 2319, Ap |
| 2 | 33 | 58.9 | 567 | 15 | US-10-270-333-126 Sequence 126, App |
| 3 | 33 | 58.9 | 622 | 10 | US-09-738-626-4919 Sequence 4919, Ap |
| 4 | 33 | 58.9 | 747 | 11 | US-09-866-050A-663 Sequence 663, App |
| 5 | 33 | 58.9 | 3472 | 12 | US-10-029-120-4 Sequence 4, Appli |
| 6 | 33 | 58.9 | 3472 | 14 | US-10-027-806-4 Sequence 4, Appli |
| 7 | 33 | 58.9 | 3472 | 14 | US-10-034-623-4 Sequence 4, Appli |
| 8 | 33 | 58.9 | 3472 | 15 | US-10-027-801-4 Sequence 4, Appli |
| 9 | 32 | 57.1 | 36 | 12 | US-10-199-820-241 Sequence 241, App |
| 10 | 32 | 57.1 | 236 | 12 | US-10-029-386-32076 Sequence 32076, A |
| 11 | 32 | 57.1 | 242 | 12 | US-10-034-749-2076 Sequence 2076, Ap |
| 12 | 32 | 57.1 | 254 | 9 | US-09-778-927A-53 Sequence 53, Appli |
| 13 | 32 | 57.1 | 276 | 12 | US-10-116-275-114 Sequence 114, App |
| 14 | 32 | 57.1 | 289 | 10 | US-09-919-497-54 Sequence 54, Appli |
| 15 | 32 | 57.1 | 289 | 14 | US-10-024-066-2 Sequence 2, Appli |

| | | | | | |
|----|------|------|-----|----|---------------------------------------|
| 16 | 32 | 57.1 | 289 | 14 | US-10-024-066-4 Sequence 4, Appli |
| 17 | 32 | 57.1 | 292 | 12 | US-10-116-275-275 Sequence 275, App |
| 18 | 32 | 57.1 | 315 | 12 | US-10-238-075-1315 Sequence 1315, Ap |
| 19 | 32 | 57.1 | 381 | 12 | US-09-769-736-24 Sequence 24, Appli |
| 20 | 32 | 57.1 | 519 | 11 | US-09-934-455-164 Sequence 164, App |
| 21 | 32 | 57.1 | 519 | 12 | US-10-302-267-36 Sequence 36, Appli |
| 22 | 32 | 57.1 | 653 | 11 | US-09-820-843A-26 Sequence 26, Appli |
| 23 | 32 | 57.1 | 715 | 11 | US-09-252-088-16 Sequence 16, Appli |
| 24 | 32 | 57.1 | 715 | 12 | US-10-340-792-16 Sequence 16, Appli |
| 25 | 32 | 57.1 | 793 | 11 | US-09-252-088-15 Sequence 15, Appli |
| 26 | 32 | 57.1 | 793 | 12 | US-10-340-792-15 Sequence 15, Appli |
| 27 | 32 | 57.1 | 822 | 12 | US-09-769-736-18 Sequence 18, Appli |
| 28 | 31.5 | 56.2 | 662 | 14 | US-10-047-542-80 Sequence 80, Appli |
| 29 | 31.5 | 56.2 | 847 | 10 | US-09-870-759-52 Sequence 52, Appli |
| 30 | 31.5 | 56.2 | 847 | 12 | US-09-751-708A-52 Sequence 52, Appli |
| 31 | 31.5 | 56.2 | 847 | 15 | US-10-207-655-133 Sequence 133, App |
| 32 | 31 | 55.4 | 59 | 10 | US-09-948-080-14 Sequence 14, Appli |
| 33 | 31 | 55.4 | 73 | 9 | US-09-864-761-40832 Sequence 40832, A |
| 34 | 31 | 55.4 | 153 | 12 | US-10-029-386-32486 Sequence 32486, A |
| 35 | 31 | 55.4 | 192 | 11 | US-09-986-480-171 Sequence 171, App |
| 36 | 31 | 55.4 | 192 | 14 | US-10-001-857-119 Sequence 119, App |
| 37 | 31 | 55.4 | 216 | 15 | US-10-156-761-12762 Sequence 12762, A |
| 38 | 31 | 55.4 | 251 | 15 | US-10-106-698-6257 Sequence 6257, Ap |
| 39 | 31 | 55.4 | 260 | 9 | US-09-815-242-13489 Sequence 13489, A |
| 40 | 31 | 55.4 | 260 | 9 | US-09-815-242-13613 Sequence 13613, A |
| 41 | 31 | 55.4 | 299 | 9 | US-09-815-242-10697 Sequence 10697, A |
| 42 | 31 | 55.4 | 336 | 11 | US-09-782-974C-86 Sequence 86, Appli |
| 43 | 31 | 55.4 | 337 | 9 | US-09-943-798-4 Sequence 4, Appli |
| 44 | 31 | 55.4 | 337 | 11 | US-09-885-453-1 Sequence 1, Appli |
| 45 | 31 | 55.4 | 337 | 12 | US-10-017-161-526 Sequence 526, App |

ALIGNMENTS

RESULT 1

US-10-094-749-2319
; Sequence 2319, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOTYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2319
; LENGTH: 273
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-2319

Query Match 58.9%; Score 33; DB 12; Length 273;
Best Local Similarity 71.4%; Pred. No. 1.7e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 5 PXGMHYS 11
|:|:|:|
Db 151 PGLHYS 157

RESULT 2

US-10-270-333-126
; Sequence 126, Application US/10270333
; Publication No. US20030092124A1
; GENERAL INFORMATION:
; APPLICANT: Cravchik, Anibal
; TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING GPCR PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF AS INSECTICIDAL TARGETS
; FILE REFERENCE: CL000733CON
; CURRENT APPLICATION NUMBER: US/10/270,333
; CURRENT FILING DATE: 2002-10-15
; PRIOR APPLICATION NUMBER: 60/169,677
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: 60/175,691
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/191,638
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 198
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 126
; LENGTH: 567
; TYPE: PRT
; ORGANISM: Drosophila
US-10-270-333-126

Query Match 58.9%; Score 33; DB 15; Length 567;
Best Local Similarity 83.3%; Pred. No. 3.7e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 PXGMHY 10
|:|:|:|
Db 402 PGMHY 407

RESULT 3

US-09-738-626-4919
; Sequence 4919, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 4919

; LENGTH: 622

; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4919

Query Match 58.9%; Score 33; DB 10; Length 622;
Best Local Similarity 50.0%; Pred. No. 4.1e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 2; Gaps 1;

Qy 1 KEVVPXGM-HY 10
|:|:|:|:|
Db 46 BEIIPGVNHY 57

RESULT 4

US-09-866-050A-663
; Sequence 663, Application US/09866050A
; Publication No. US20030040471A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods for Their Use
; FILE REFERENCE: 11000.1011c4U
; CURRENT APPLICATION NUMBER: US/09/866,050A
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 663
; LENGTH: 747
; TYPE: PRT
; ORGANISM: Rat
US-09-866-050A-663

Query Match 58.9%; Score 33; DB 11; Length 747;
Best Local Similarity 71.4%; Pred. No. 5e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 5 PXGMHYS 11
|:|:|:|
Db 627 PGLHYS 633

RESULT 5

US-10-029-120-4
; Sequence 4, Application US/10029120
; Publication No. US20030175708A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCOEP.002A
; CURRENT APPLICATION NUMBER: US/10/029,120
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-029-120-4

Query Match 58.9%; Score 33; DB 12; Length 3472;
Best Local Similarity 45.5%; Pred. No. 2.6e+03;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

```

Qy      1 BEVVPXGMHYS 11
      ||:||:||:
Db      2294 EDVIPRGISFS 2304

RESULT 6
US-10-027-806-4
; Sequence 4, Application US/10027806
; Publication No. US20020160476A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCCRP.002A
; CURRENT APPLICATION NUMBER: US/10/027,806
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-027-806-4

Query Match      58.9%; Score 33; DB 14; Length 3472;
Best Local Similarity 45.5%; Pred. No. 2.6e+03;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy      1 BEVVPXGMHYS 11
      ||:||:||:
Db      2294 EDVIPRGISFS 2304

RESULT 7
US-10-034-623-4
; Sequence 4, Application US/10034623
; Publication No. US20020198365A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCCRP.002A
; CURRENT APPLICATION NUMBER: US/10/034,623
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: 60/102,294
; PRIOR FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-034-623-4

Query Match      58.9%; Score 33; DB 14; Length 3472;
Best Local Similarity 45.5%; Pred. No. 2.6e+03;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy      1 BEVVPXGMHYS 11
      ||:||:||:
Db      2294 EDVIPRGISFS 2304

RESULT 8
US-10-027-801-4
; Sequence 4, Application US/10027801
; Publication No. US20030054364A1

```

```

; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCCRP.002A
; CURRENT APPLICATION NUMBER: US/10/027,801
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-027-801-4

Query Match      58.9%; Score 33; DB 15; Length 3472;
Best Local Similarity 45.5%; Pred. No. 2.6e+03;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy      1 BEVVPXGMHYS 11
      ||:||:||:
Db      2294 EDVIPRGISFS 2304

RESULT 9
US-10-199-820-241
; Sequence 241, Application US/10199820
; Publication No. US20030180739A1
; GENERAL INFORMATION:
; APPLICANT: Board of Trustees of the University of Illinois
; APPLICANT: Primiano, Thomas
; APPLICANT: Chang, Bey-dih
; APPLICANT: Roninson, Igor
; TITLE OF INVENTION: Methods and Reagents for Identifying Gene Targets for Treating
; FILE REFERENCE: 99,216-U
; CURRENT APPLICATION NUMBER: US/10/199,820
; CURRENT FILING DATE: 2002-09-23
; NUMBER OF SEQ ID NOS: 314
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 241
; LENGTH: 36
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-199-820-241

Query Match      57.1%; Score 32; DB 12; Length 36;
Best Local Similarity 60.0%; Pred. No. 31;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy      1 BEVVPXGMHY 10
      ||:||:||:
Db      20 BEVPFLAMNY 29

RESULT 10
US-10-029-386-32076
; Sequence 32076, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 32076

```

; LENGTH: 236
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: MAP TO Z98050.1
 ; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.5
 ; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
 ; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.6
 ; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.9
 ; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.3
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.5
 ; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.3
 ; OTHER INFORMATION: SWISSPROT HIT: P15822, EVALUE 1.00e-125
 US-10-029-386-32076

Query Match 57.1%; Score 32; DB 12; Length 236;
 Best Local Similarity 66.7%; Pred. No. 2.3e+02;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 3 VVPXGMHY 11
 Db 80 VVPAGLTYS 88

RESULT 11

US-10-094-749-2076
 ; Sequence 2076, Application US/10094749
 ; Publication No. US20030219741A1

; GENERAL INFORMATION:
 ; APPLICANT: ISOGAI, TAKAO
 ; APPLICANT: SUGIYAMA, TOMOYASU
 ; APPLICANT: OTSUKI, TETSUJI
 ; APPLICANT: WAKAMATSU, AI
 ; APPLICANT: SATO, HIROYUKI
 ; APPLICANT: ISHII, SHIZUKO
 ; APPLICANT: YAMAMOTO, JUN-ICHI
 ; APPLICANT: ISONO, YUUKO
 ; APPLICANT: HIO, YURI
 ; APPLICANT: OTSUKA, KAORU
 ; APPLICANT: NAGAI, KEIICHI
 ; APPLICANT: IRIE, RYOTARO
 ; APPLICANT: TAMECHIKA, ICHIRO
 ; APPLICANT: SEKI, NAOHICO
 ; APPLICANT: YOSHIKAWA, TSUTOMU
 ; APPLICANT: OTSUKA, MOTOTYUKI
 ; APPLICANT: NAGAHARI, KENJI
 ; APPLICANT: MASUHO, YASUHIKO
 ; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
 ; FILE REFERENCE: 084335/0160
 ; CURRENT APPLICATION NUMBER: US/10/094,749
 ; CURRENT FILING DATE: 2002-03-12
 ; PRIOR APPLICATION NUMBER: 60/350,435
 ; PRIOR FILING DATE: 2002-01-24
 ; PRIOR APPLICATION NUMBER: JP 2001-328381
 ; PRIOR FILING DATE: 2001-09-14
 ; NUMBER OF SEQ ID NOS: 3381
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 2076
 ; LENGTH: 242
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens

US-10-094-749-2076

Query Match 57.1%; Score 32; DB 12; Length 242;
 Best Local Similarity 60.0%; Pred. No. 2.4e+02;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVVPXGMHY 10
 Db 25 BEVFPPLAMNY 34

RESULT 12

US-09-778-927A-53
 ; Sequence 53, Application US/09778927A
 ; Patent No. US20020068342A1

; GENERAL INFORMATION:
 ; APPLICANT: KHOSRAVI, Rami et al.
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACID AND AMINO ACID SEQUENCES AND NOVEL
 ; FILE REFERENCE: 2786-0150P
 ; CURRENT APPLICATION NUMBER: US/09/778,927A
 ; CURRENT FILING DATE: 2001-02-08
 ; PRIOR APPLICATION NUMBER: IL 134453
 ; PRIOR FILING DATE: 2000-02-09
 ; PRIOR APPLICATION NUMBER: IL135341
 ; PRIOR FILING DATE: 2000-03-29
 ; NUMBER OF SEQ ID NOS: 81
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 53
 ; LENGTH: 254
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)..(254)
 ; OTHER INFORMATION: Xaa = any amino acid, unknown, or other

US-09-778-927A-53

Query Match 57.1%; Score 32; DB 9; Length 254;
 Best Local Similarity 60.0%; Pred. No. 2.5e+02;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVVPXGMHY 10
 Db 74 BEVFPPLAMNY 83

RESULT 13

US-10-116-275-114
 ; Sequence 114, Application US/10116275
 ; Publication No. US20030211476A1

; GENERAL INFORMATION:
 ; APPLICANT: Elan Pharmaceutical Technology
 ; APPLICANT: O'Mahony, Daniel J.
 ; APPLICANT: Brayden, David
 ; APPLICANT: Byrne, Daragh
 ; APPLICANT: Lambkin, Imelda
 ; APPLICANT: Higgins, Lisa
 ; TITLE OF INVENTION: Genetic Analysis of Peyer's Patches and M Cells and Methods and
 ; FILE REFERENCE: E1067/20087
 ; CURRENT APPLICATION NUMBER: US/10/116,275
 ; CURRENT FILING DATE: 2002-10-04
 ; NUMBER OF SEQ ID NOS: 349
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 114
 ; LENGTH: 276
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens

US-10-116-275-114

Query Match 57.1%; Score 32; DB 12; Length 276;
 Best Local Similarity 60.0%; Pred. No. 2.7e+02;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVVPXGMHY 10
 Db 75 BEVFPPLAMNY 84

RESULT 14

US-09-919-497-54
 ; Sequence 54, Application US/09919497
 ; Patent No. US20020106662A1

; GENERAL INFORMATION:

;; APPLICANT: Mutter, George L.
;; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
;; FILE REFERENCE: B0801/7225
;; CURRENT APPLICATION NUMBER: US/09/919,497
;; CURRENT FILING DATE: 2001-07-31
;; PRIOR APPLICATION NUMBER: US 60/221,735
;; PRIOR FILING DATE: 2000-07-31
;; NUMBER OF SEQ ID NOS: 100
;; SOFTWARE: PatentIn version 3.0
;; SEQ ID NO 54
;; LENGTH: 289
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-09-919-497-54

Query Match 57.1%; Score 32; DB 10; Length 289;
Best Local Similarity 60.0%; Pred. No. 2.9e+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMHY 10
||| | | | | | | | | |
Db 74 BEVFPPLANNY 83

RESULT 15
US-10-024-066-2
; Sequence 2, Application US/10024066
; Publication No. US20020166134A1
; GENERAL INFORMATION:
; APPLICANT: Field, Loren J.
; APPLICANT: Pasumathi, Kishore Babu S.
; TITLE OF INVENTION: CARDIOMYOCYTES WITH ENHANCED PROLIFERATIVE POTENTIAL,
; TITLE OF INVENTION: AND METHODS FOR PREPARING AND USING SAME
; FILE REFERENCE: 7037-450
; CURRENT APPLICATION NUMBER: US/10/024,066
; CURRENT FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: 60/139,942
; PRIOR FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: PCT/US00/16827
; PRIOR FILING DATE: 2000-06-19
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 289
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-024-066-2

Query Match 57.1%; Score 32; DB 14; Length 289;
Best Local Similarity 60.0%; Pred. No. 2.9e+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMHY 10
||| | | | | | | | | |
Db 73 BEVFPPLANNY 82

Search completed: December 22, 2003, 17:32:43
Job time : 21.9333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 17:24:36 ; Search time 9.06667 Seconds
(without alignments)
116.675 Million cell updates/sec

Title: US-09-909-164-11
Perfect score: 56
Sequence: 1 BEVDPXGMHYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 76:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query % | Match | Length | DB | ID | Description |
|------------|-------|---------|-------|--------|--------|----|---------------------|
| 1 | 37 | 66.1 | 308 | 2 | A72207 | | ftsH proteinase ac |
| 2 | 37 | 66.1 | 1057 | 2 | F89892 | | carbamoyl-phosphat |
| 3 | 36 | 64.3 | 102 | 2 | A42452 | | V1 protein - tobas |
| 4 | 36 | 64.3 | 252 | 2 | AE2001 | | hypothetical prote |
| 5 | 36 | 64.3 | 460 | 2 | S69046 | | hypothetical prote |
| 6 | 36 | 64.3 | 743 | 2 | S38143 | | hypothetical prote |
| 7 | 35 | 62.5 | 156 | 2 | D82618 | | conserved hypothet |
| 8 | 35 | 62.5 | 233 | 2 | T02590 | | DNA binding protei |
| 9 | 35 | 62.5 | 311 | 2 | H69194 | | GMP synthetase, su |
| 10 | 35 | 62.5 | 425 | 2 | T24111 | | hypothetical prote |
| 11 | 35 | 62.5 | 510 | 2 | G86430 | | T518.1 protein - A |
| 12 | 34 | 60.7 | 264 | 2 | G69117 | | diphthine synthase |
| 13 | 34 | 60.7 | 279 | 2 | C75538 | | hypothetical prote |
| 14 | 34 | 60.7 | 350 | 2 | B75478 | | 3-dehydroquinatase |
| 15 | 34 | 60.7 | 355 | 2 | T35025 | | probable DNA ligase |
| 16 | 34 | 60.7 | 360 | 2 | E69086 | | cell division prot |
| 17 | 34 | 60.7 | 425 | 2 | E83903 | | hypothetical prote |
| 18 | 34 | 60.7 | 426 | 2 | S58132 | | Sle1 protein precu |
| 19 | 34 | 60.7 | 495 | 2 | T28717 | | hypothetical prote |
| 20 | 34 | 60.7 | 1028 | 2 | AF3286 | | ATP-dependent DNA |
| 21 | 33 | 58.9 | 156 | 2 | S54619 | | hypothetical prote |
| 22 | 33 | 58.9 | 367 | 2 | E83607 | | polyamine transpor |
| 23 | 33 | 58.9 | 441 | 2 | G82253 | | conserved hypothet |
| 24 | 33 | 58.9 | 466 | 2 | G71542 | | probable amino aci |
| 25 | 33 | 58.9 | 466 | 2 | H81697 | | amino acid antipor |
| 26 | 33 | 58.9 | 487 | 2 | S65811 | | finger protein (cl |
| 27 | 33 | 58.9 | 514 | 1 | HQDVLB | | cytochrome-c3 hydr |
| 28 | 33 | 58.9 | 534 | 2 | A69284 | | coenzyme F420-quin |
| 29 | 33 | 58.9 | 545 | 2 | T08564 | | hypothetical prote |

| | | | | | | |
|----|----|------|------|---|--------|---------------------|
| 30 | 33 | 58.9 | 627 | 2 | A69663 | DNA mismatch repair |
| 31 | 33 | 58.9 | 716 | 1 | JC5061 | macrophage-stimula |
| 32 | 33 | 58.9 | 1257 | 2 | S44754 | C14B9.8 protein - |
| 33 | 33 | 58.9 | 1396 | 2 | S36851 | L-shaped tail fibe |
| 34 | 33 | 58.9 | 3472 | 2 | T31308 | hypothetical 367K |
| 35 | 32 | 57.1 | 126 | 2 | A25905 | tachykinin B precu |
| 36 | 32 | 57.1 | 197 | 2 | D71640 | heme exporter prot |
| 37 | 32 | 57.1 | 225 | 2 | S57810 | hypothetical prote |
| 38 | 32 | 57.1 | 233 | 2 | E97120 | ribosomal protein |
| 39 | 32 | 57.1 | 267 | 2 | T07215 | probable Thua prot |
| 40 | 32 | 57.1 | 270 | 2 | C95881 | hypothetical prote |
| 41 | 32 | 57.1 | 283 | 2 | T25737 | cyclin D2 - rat |
| 42 | 32 | 57.1 | 288 | 2 | JC4011 | cyclin D2 - rat |
| 43 | 32 | 57.1 | 288 | 2 | I58372 | cyclin D2 - mouse |
| 44 | 32 | 57.1 | 289 | 2 | A41984 | cyclin D2 - human |
| 45 | 32 | 57.1 | 289 | 2 | A42822 | |

ALIGNMENTS

RESULT 1

A72207 ftsH proteinase activity modulator HflK - Thermotoga maritima (strain MSB8)

C:Species: Thermotoga maritima

C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000

C:Accession: A72207

R:Neilson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hick
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, I
C.M.

Nature 399, 323-329, 1999

A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome s
A:Reference number: A72200; MUID:99287316; PMID:10360571

A:Accession: A72207

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-308 <ARN>

A:Cross-references: GB:AS001819; GB:AE000512; NID:G4982396; PIDN:AAD36885.1; PID:G4982

A:Experimental source: strain MSB8

C:Genetics:

A:Gene: TM1822

C:Superfamily: erythrocyte band 7 integral membrane protein

Query Match 66.1%; Score 37; DB 2; Length 308;
Best Local Similarity 75.0%; Pred. No. 10;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMHY 10

DB 41 VVPSGIHY 48

RESULT 2

F89892 carbamoyl-phosphate synthase large chain [imported] - Staphylococcus aureus (strain N3

C:Species: Staphylococcus aureus

C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001

C:Accession: F89892

R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Og
ma, A.; Mizutani-U, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramoto, K.

Lancet 357, 1225-1240, 2001

A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A:Reference number: A89758; MUID:21311952; PMID:11418146

A:Accession: F89892

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1057 <KUR>

A:Cross-references: GB:BA000018; PID:gl3701002; PIDN:BA842298.1; GSPDB:GN00149

A:Experimental source: strain N315

C:Genetics:

A:Gene: pyrAB

C:Superfamily: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain; bioti

Query Match 66.1%; Score 37; DB 2; Length 1057;
 Best Local Similarity 60.0%; Pred. No. 39;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 2 EVVPXGMHYS 11
 :|||:|:
 DB 190 EIVSNGLHYS 199

RESULT 3

A42452
 V1 protein - tobacco yellow dwarf virus (strain Australia)
 C:Species: tobacco yellow dwarf virus
 C>Date: 15-Jan-1993 #sequence_revision 15-Jan-1993 #text_change 08-Oct-1999
 C:Accession: A42452
 R:Morris, B.A.M.; Richardson, K.A.; Haley, A.; Zhan, X.; Thomas, J.E.
 A:Title: The nucleotide sequence of the infectious cloned DNA component of tobacco yellow
 A:Reference number: A42452; MUID:92188538; PMID:1546458
 A:Accession: A42452
 A:Molecule type: DNA
 A:Residues: 1-102 <MOR>
 A:CROSS-references: GB:M81103; NID:g335283; PIDN:AAA47947.1; PID:g335284

Query Match 64.3%; Score 36; DB 2; Length 102;
 Best Local Similarity 60.0%; Pred. No. 5;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 QY 2 EVVPXGMHYS 11
 :|||:|:
 DB 7 QVVPSGINYS 16

RESULT 4

AE2001
 hypothetical protein alr1563 [imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp. PCC 7120
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
 C:Accession: AE2001
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriiguchi,
 N.; Kaneko, T.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A:Title: Complete genomic sequence of the filamentous nitrogen-fixing Cyanobacterium Anabaena
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AE2001
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-252 <KUR>
 A:CROSS-references: GB:BA000019; PIDN:BA077929.1; PID:g17135383; GSPDB:GN00179
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: alr1563

Query Match 64.3%; Score 36; DB 2; Length 252;
 Best Local Similarity 50.0%; Pred. No. 13;
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 QY 1 EBVPXGMHYS 10
 :|||:|:
 DB 235 EMIVPAGLHF 244

RESULT 5

S69046
 hypothetical protein YPL139C - yeast (Saccharomyces cerevisiae)
 C:Species: Saccharomyces cerevisiae
 C>Date: 22-Aug-1996 #sequence_revision 06-Sep-1996 #text_change 17-Mar-2000
 C:Accession: S69046
 R:Hall, J.; DePaulo, T.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.; Storms, R.K.;
 submitted to the EMBL Data Library, December 1995
 A:Description: The sequence of Saccharomyces cerevisiae chromosome XVI left arm.

A:Reference number: S69040
 A:Accession: S69046
 A:Molecule type: DNA
 A:Residues: 1-460 <HAL>
 A:CROSS-references: EMBL:U43703; NID:gl244769; PIDN:AAB68221.1; PID:gl244776; MIPS:YPL
 C:Genetics:
 A:Gene: SGD:UMB1
 A:CROSS-references: SGD:S0006060; MIPS:YPL139C
 A:Map position: 16L
 C:Superfamily: Saccharomyces cerevisiae transcription modulator WTMI

Query Match 64.3%; Score 36; DB 2; Length 460;
 Best Local Similarity 62.5%; Pred. No. 25;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 3 VVPXGMHY 10
 :|||:|:
 DB 85 IVPLGLHY 92

RESULT 6

S38143
 hypothetical protein YBL011w homolog YKR067w - yeast (Saccharomyces cerevisiae)
 C:Species: Saccharomyces cerevisiae
 C>Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 19-Apr-2002
 C:Accession: S38143
 R:van Vliet-Reedijk, J.C.; Planta, R.J.
 submitted to the Protein Sequence Database, March 1994
 A:Reference number: S38130
 A:Accession: S38143
 A:Molecule type: DNA
 A:Residues: 1-743 <VAN>
 A:CROSS-references: EMBL:Z28292; NID:g486536; PIDN:CAA82146.1; PID:g486537; MIPS:YKR06
 A:Experimental source: strain S288C
 C:Genetics:
 A:Gene: SGD:GPT2
 A:CROSS-references: SGD:S0001775
 A:Map position: 11R
 C:Keywords: transmembrane protein

Query Match 64.3%; Score 36; DB 2; Length 743;
 Best Local Similarity 75.0%; Pred. No. 43;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMHY 10
 :|||:|:
 DB 294 VVPCGLHY 301

RESULT 7

DB2618
 conserved hypothetical protein XF1950 [imported] - Xylella fastidiosa (strain 9a5c)
 C:Species: Xylella fastidiosa
 C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
 C:Accession: DB2618
 R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing
 Nature 406, 151-157, 2000
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A:Reference number: AB2515; MUID:20365717; PMID:10910347
 A:Note: for a complete list of authors see reference number A59328 below
 A:Accession: DB2618
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-156 <SIM>
 A:CROSS-references: GB:AE004014; GB:AE003849; NID:g9107044; PIDN:AAF84752.1; GSPDB:GNO
 A:Experimental source: strain 9a5c
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, P.A.; Acencio, M.; Alvarenga, R.;
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer,
 as-Neto, E.; Docena, C.; El-Dorry, H.; Facincan, A.P.; Ferreira, A.J.S.
 submitted to GenBank, June 2000
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fro
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Lai
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins,

A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
 A:Authors: da Silva, A.C.R.; da Silva, P.R.; da Silva, A.M.; Silva Jr., M.A.; da Silveir
 M.; Tsuchiko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.;
 A:Reference number: A59328
 A:Contents: annotation
 C:Genetics:
 A:Gene: XF1950

Query Match 62.5%; Score 35; DB 2; Length 156;
 Best Local Similarity 55.6%; Pred. No. 13;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXGMH 9

DB 119 BEILPQGVH 127

RESULT 8

DNA binding protein EREBP-2 - common tobacco
 T02590
 C:Species: Nicotiana tabacum (common tobacco)
 C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 21-Jul-2000
 C:Accession: T02590
 R:Ohme-Takagi, M.; Shinshi, H.
 Plant Cell 7, 173-182, 1995
 A:Title: Ethylene-inducible DNA binding proteins that interact with an ethylene responsi
 A:Reference number: Z14671; MUID:95276459; PMID:7756828
 A:Accession: T02590
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-233 <OHM>
 A:Cross-references: EMBL:D38126; NID:G790362; PIDN:BAA07324.1; PID:G1208498
 A:Experimental source: strain BY4, tissue-type leaf

Query Match 62.5%; Score 35; DB 2; Length 233;
 Best Local Similarity 60.0%; Pred. No. 19;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMHY 10

DB 90 QAVPQGRHY 99

RESULT 9

GMP synthetase, subunit B - Methanobacterium thermoautotrophicum (strain Delta H)
 H69194
 C:Species: Methanobacterium thermoautotrophicum
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
 C:Accession: H69194
 R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
 Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwan, N.
 K.; S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
 J. Bacteriol. 179, 7135-7155, 1997
 A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
 A:Reference number: A69000; MUID:98037514; PMID:9371463
 A:Accession: H69194
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-311 <MTH>
 A:Cross-references: GB:AE000850; GB:AE000666; NID:G2621794; PIDN:AAB85215.1; PID:G262179

A:Experimental source: strain Delta H

C:Genetics:

A:Gene: MTH710

A:Start codon: GTG

Query Match 62.5%; Score 35; DB 2; Length 311;
 Best Local Similarity 63.6%; Pred. No. 27;
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMHYS 11

||||| :|:

DB 219 BEVVEGLHES 229

RESULT 10

T24111

Hypothetical protein R10D12.10 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T24111

R:Percy, C.

submitted to the EMBL Data Library, October 1996

A:Reference number: Z19842

A:Accession: T24111

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-425 <WIL>

A:Cross-references: EMBL:Z81109; PIDN:CAB03241.1; GSPDB:GN00023; CESP:R10D12.10

A:Experimental source: clone R10D12

C:Genetics:

A:Gene: CESP:R10D12.10

A:Map position: 5

A:Introns: 23/3; 56/3; 113/3; 257/2

Query Match 62.5%; Score 35; DB 2; Length 425;

Best Local Similarity 50.0%; Pred. No. 37;

Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMHY 10

||||| :|:

DB 335 E01VPGGLQY 344

RESULT 11

GS6430

T518.1 protein - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 17-May-2002

C:Accession: GS6430

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alon

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, T.H.; Dewar, X.

ansen, N.P.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Mafti, R.; Marzial

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: AB6141; MUID:21016719; PMID:11130712

A:Accession: GS6430

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-510 <STO>

A:Cross-references: GB:AB005172; NID:94587512; PIDN:AAD25743.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

C:Superfamily: hexose phosphate transport protein uhpt

Query Match 62.5%; Score 35; DB 2; Length 510;

Best Local Similarity 60.0%; Pred. No. 45;

Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMHY 10

||||| :|:

DB 12 BEVXPPIHF 21

RESULT 12

GS9117

diphthine synthase - Methanobacterium thermoautotrophicum (strain Delta H)

C:Species: Methanobacterium thermoautotrophicum

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 17-Feb-2003

C:Accession: G69117

R.; Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Oiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiواني, N.; S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
 J. Bacteriol. 179, 7135-7155, 1997
 A>Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: func0
 A:Reference number: A69000; MUID:98037514; PMID:93711463
 A:Accession: G69117
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-264 <MTH>
 A:Cross-references: GB:AE000940; GB:AE000666; NID:G2623011; PID:AA86340.1; PID:G262301
 A:Experimental source: strain Delta H
 C:Genetics:
 A:Gene: MTH1874

Query Match 60.7%; Score 34; DB 2; Length 264;
 Best Local Similarity 62.5%; Pred. No. 35;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMHY 10
 |||||
 Db 235 VVPAGLHF 242

RESULT 13

C75538
 hypothetical protein - Deinococcus radiodurans (strain R1)
 C:Species: Deinococcus radiodurans
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 28-Jul-2000
 C:Accession: C75538
 R.White, O.; Eiseen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
 M.; Shen, M.; Vanathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
 Science 286, 1571-1577, 1999
 A>Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A:Reference number: A75250; MUID:20036896; PMID:10567266
 A:Accession: C75538
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-279 <WHI>
 A:Cross-references: GB:AE001889; GB:AE000513; NID:G6457944; PID:AAF09867.1; PID:G645795
 A:Experimental source: strain R1
 C:Genetics:
 A:Gene: DR0271
 A:Map position: 1
 C:Superfamily: Deinococcus radiodurans hypothetical protein DR0271

Query Match 60.7%; Score 34; DB 2; Length 279;
 Best Local Similarity 75.0%; Pred. No. 38;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 VVPXGMHY 11
 |||||
 Db 100 VPLGRHY 107

RESULT 14

B75478
 3-dehydroquininate synthase - Deinococcus radiodurans (strain R1)
 C:Species: Deinococcus radiodurans
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
 C:Accession: B75478
 R.White, O.; Eiseen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
 M.; Shen, M.; Vanathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
 Science 286, 1571-1577, 1999
 A>Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A:Reference number: A75250; MUID:20036896; PMID:10567266
 A:Accession: B75478
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-350 <WHI>
 A:Cross-references: GB:AE001932; GB:AE000513; NID:G6458481; PID:AAF10353.1; PID:G645848

A:Experimental source: strain R1
 C:Genetics:
 A:Gene: DR0777
 A:Map position: 1
 C:Superfamily: 3-dehydroquininate synthase; 3-dehydroquininate synthase homology

Query Match 60.7%; Score 34; DB 2; Length 350;
 Best Local Similarity 60.0%; Pred. No. 48;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EVVPXGMHY 11
 |||||
 Db 252 EAVAYGMHYA 261

RESULT 15

T35025
 probable DNA ligase - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
 C:Accession: T35025
 R.Seeger, S.; Harris, D.; James, K.D.; Parkhill, J.; Barrall, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, June 1999
 A:Reference number: Z21565
 A:Accession: T35025
 A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
 A:Residues: 1-355 <SEE>
 A:Cross-references: EMBL:AL079355; PIDN:CAB45581.1; GSPDB:GN000070; SCORDB:SC4C6.17c
 A:Experimental source: strain A3(2)
 C:Genetics:
 A:Gene: SCORDB:SC4C6.17c

Query Match 60.7%; Score 34; DB 2; Length 355;
 Best Local Similarity 71.4%; Pred. No. 49;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 VVPXGMHY 10
 :|||
 Db 20 IPPGMHY 26

Search completed: December 22, 2003, 17:44:57
 Job time : 9.06667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:43:51 ; Search time 4.6 Seconds
(without alignments)
112.455 Million cell updates/sec

Title: US-09-909-164-11
Perfect score: 56
Sequence: 1 BEVVPXGMHYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 38 | 67.9 | 1058 | 1 | Q8rg86 fusobacteri |
| 2 | 37 | 66.1 | 426 | 1 | Q8rqx9 vibrio para |
| 3 | 37 | 66.1 | 1057 | 1 | Q99ur5 staphylococ |
| 4 | 37 | 66.1 | 1057 | 1 | P59940 staphylococ |
| 5 | 36 | 64.3 | 102 | 1 | P31619 tobacco yel |
| 6 | 36 | 64.3 | 460 | 1 | Q03010 saccharomyc |
| 7 | 36 | 64.3 | 743 | 1 | YK47 YEAST |
| 8 | 35 | 62.5 | 227 | 1 | ID11 MESAU |
| 9 | 35 | 62.5 | 308 | 1 | GAAB METH |
| 10 | 34 | 60.7 | 426 | 1 | SL51 YARLI |
| 11 | 33 | 58.9 | 441 | 1 | YL15 VIBPA |
| 12 | 33 | 58.9 | 513 | 1 | PHSL DESBA |
| 13 | 33 | 58.9 | 627 | 1 | MUTL BACSU |
| 14 | 33 | 58.9 | 1188 | 1 | KPBA CAEEL |
| 15 | 33 | 58.9 | 1396 | 1 | VLTFF BPTS |
| 16 | 32 | 57.1 | 126 | 1 | TKNK BOVIN |
| 17 | 32 | 57.1 | 233 | 1 | RS2 CLOAB |
| 18 | 32 | 57.1 | 267 | 1 | RR2 CHLVU |
| 19 | 32 | 57.1 | 288 | 1 | CGD2 RAT |
| 20 | 32 | 57.1 | 289 | 1 | CGD2 HUMAN |
| 21 | 32 | 57.1 | 289 | 1 | CGD2 MOUSE |
| 22 | 32 | 57.1 | 291 | 1 | CGD1 BRARE |
| 23 | 32 | 57.1 | 291 | 1 | CGD1 XENLA |
| 24 | 32 | 57.1 | 291 | 1 | CGD2 CHICK |
| 25 | 32 | 57.1 | 291 | 1 | CGD2 XENLA |
| 26 | 32 | 57.1 | 292 | 1 | CGD1 CHICK |
| 27 | 32 | 57.1 | 292 | 1 | CGD3 HUMAN |
| 28 | 32 | 57.1 | 295 | 1 | CGD1 HUMAN |
| 29 | 32 | 57.1 | 295 | 1 | CGD1 MOUSE |
| 30 | 32 | 57.1 | 295 | 1 | CGD1 RAT |
| 31 | 32 | 57.1 | 341 | 1 | HYPE AZOVI |
| 32 | 32 | 57.1 | 353 | 1 | T2BA BACAR |
| 33 | 32 | 57.1 | 573 | 1 | SUOX DROME |

RESULT 1

| CARB_FUSNN | STANDARD; | PRT; | 1058 AA. |
|------------|--|------|----------|
| AC | Q8RG86; | | |
| DT | 28-FEB-2003 (Rel. 41, Created) | | |
| DT | 28-FEB-2003 (Rel. 41, Last sequence update) | | |
| DT | 28-FEB-2003 (Rel. 41, Last annotation update) | | |
| DE | Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-phosphate synthetase ammonia chain) | | |
| GN | CARB OR FN0422. | | |
| OS | Fusobacterium nucleatum (subsp. nucleatum). | | |
| OC | Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae; | | |
| OC | Fusobacterium. | | |
| OX | NCBI_TaxID=76856; | | |
| RN | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RC | STRAIN=ATCC 25586; | | |
| RX | MEDLINE=21886394; PubMed=11889109; | | |
| RA | Kapatral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A., Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L., Vasileva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A., Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R., Fonstein M., Kyprides N., Overbeek R.; | | |
| RT | "Genome sequence and analysis of the oral bacterium Fusobacterium nucleatum strain ATCC 25586." | | |
| RL | J. Bacteriol. 184:2005-2018(2002). | | |
| CC | -1- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP + phosphate + L-glutamate + carbamoyl phosphate. | | |
| CC | -1- COPACTOR: Binds 3 manganese ions per subunit (By similarity). | | |
| CC | -1- PATHWAY: Arginine biosynthesis. | | |
| CC | -1- SUBUNIT: Pyrimidine biosynthesis; first step. | | |
| CC | -1- SUBUNIT: Composed of two chains; the small (or glutamine) chain promotes the hydrolysis of glutamine to ammonia, which is used by the large (or ammonia) chain to synthesize carbamoyl phosphate (By similarity). | | |
| CC | -1- SIMILARITY: BELONGS TO THE CARB FAMILY. | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch). | | |
| CC | EMBL; AB010554; AAL94625.1; ALT_INIT. | | |
| DR | HAWAP; MF_01210; -1 | | |
| DR | InterPro; IPR006275; CarA_L_glu. | | |
| DR | InterPro; IPR005483; Cpsae_L. | | |
| DR | InterPro; IPR005479; Cpsae_L_D2. | | |
| DR | InterPro; IPR005480; Cpsae_L_D3. | | |
| DR | InterPro; IPR005481; Cpsae_L_N. | | |
| DR | InterPro; IPR004362; MGS_like. | | |
| DR | Pfam; PF00289; Cpsase_L_chain; 2. | | |
| DR | Pfam; PF02786; Cpsase_L_D2; 2. | | |
| DR | Pfam; PF02787; Cpsase_L_D3; 1. | | |

Q89a96 buchnera ap
P32784 saccharomyc
O74377 schizosacch
Q9kv29 vibrio chol
P15822 homo sapien
P20273 homo sapien
P17280 chimpanzee
P97885 rattus norv
Q58349 methanococc
P10941 cryphonetr
O28354 archaeoglob
P52210 ovis aries

ALIGNMENTS

pfam: PF02142; MGS: 1.
 DR PRINTS: PR00098; CPASASE
 DR TIGRPFams; TIGR01369; CPASaseII_lrg; 1.
 DR PROSITE; PS00866; CPASASE 1; 2.
 DR PROSITE; PS00867; CPASASE 2; 2.
 KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
 KW ATP-binding; Manganese; Complete proteome.
 FT DOMAIN 1 401
 FT DOMAIN 402 546
 FT DOMAIN 547 929
 FT DOMAIN 930 1058
 FT REPEAT 1 546
 FT REPEAT 547 1058
 FT NP_BIND 153 210
 FT NP_BIND 302 352
 FT METAL 284 298
 FT METAL 298 298
 FT METAL 300 300
 FT METAL 820 820
 FT METAL 832 832
 SQ SEQUENCE 1058 AA; 117451 MW; ED7037AF77C1E39F CRC64;
 Query Match 67.9%; Score 38; DB 1; Length 1058;
 Best Local Similarity 60.0%; Pred. No. 10;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 QY 2 EVVPXGMHYS 11
 DB 190 EIVPGLNYS 199

RESULT 2
 ID AROA VIBPA STANDARD; PRT; 426 AA.
 AC Q87QX9;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE 3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19) (5-
 enolpyruvylshikimate-3-phosphate synthase) (EPSP synthase) (EPSPS).
 GN AROA OR VP1020.
 OS Vibrio parahaemolyticus.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 OX NCBI_TaxID=670;
 RN [1]
 SEQUENCE FROM N.A.
 RC STRAIN=RIMD 2210633 / Serotype O3:K6;
 RX MEDLINE=22508454; PubMed=12620739;
 RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
 RA Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
 RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
 RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
 distinct from that of V. cholerae";
 RL Lancet 361:743-749(2003).
 CC -|- CATALYTIC ACTIVITY: Phosphoenolpyruvate + 3-phosphoshikimate =
 CC phosphate + 5-O-(1-carboxyvinyl)-3-phosphoshikimate.
 CC -|- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
 CC sixth step.
 CC -|- SUBUNIT: Monomer (By similarity).
 CC -|- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -|- SIMILARITY: Belongs to the EPSP synthase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AP005076; BAC59283.1; -;
 CC HAMAP; MF_00210; -; 1.

pfam: PF00104; EPSP SYNTHASE 1; 1.
 DR PROSITE; PS00885; EPSP SYNTHASE 2; 1.
 KW Aromatic amino acid biosynthesis; Transferase; Complete proteome.
 SQ SEQUENCE 426 AA; 46094 MW; 373D39CC5BA1F70F CRC64;
 Query Match 66.1%; Score 37; DB 1; Length 426;
 Best Local Similarity 60.0%; Pred. No. 6.3;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 1 BEVVPXGMHY 10
 DB 223 BEVVPAGHY 232

RESULT 3
 ID CARB STAM STANDARD; PRT; 1057 AA.
 AC Q99UR5;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
 phosphate synthetase ammonia chain).
 GN CARB OR PYRAB OR SAV1203 OR SA1046.
 OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
 CC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 OX NCBI_TaxID=158878; 158879;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Mu50 / ATCC 700699, and N315;
 RX MEDLINE=21311952; PubMed=11418146;
 RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
 RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
 RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
 RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
 RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
 RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
 RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.,
 RT "Whole genome sequencing of methicillin-resistant Staphylococcus
 aureus";
 RL Lancet 357:1225-1240(2001).
 CC -|- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
 CC phosphate + L-glutamate + carbamoyl phosphate.
 CC -|- COPACTOR: Binds 3 manganese ions per subunit (By similarity).
 CC -|- PATHWAY: Arginine biosynthesis.
 CC -|- PATHWAY: Pyrimidine biosynthesis; first step.
 CC -|- SUBUNIT: Composed of two chains; the small (or glutamine) chain
 CC promotes the hydrolysis of glutamine to ammonia, which is used by
 CC the large (or ammonia) chain to synthesize carbamoyl phosphate (By
 CC similarity).
 CC -|- SIMILARITY: BELONGS TO THE CARB FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AP003361; BAB57365.1; -;
 CC EMBL; AP003132; BAB42298.1; -;
 CC FIR; F89892; F89892.
 CC HAMAP; MF_01210; -; 1.
 CC InterPro; IPR006275; CarA_L_glu.
 CC InterPro; IPR005483; CPase_L.
 CC InterPro; IPR005479; CPase_L_D2.
 CC InterPro; IPR005480; CPase_L_D3.
 CC InterPro; IPR005481; CPase_L_N.
 CC InterPro; IPR004362; MGS like.
 CC Pfam; PF00289; CPase_L_Chain; 2.

```

DR Pfam; PF02786; CPase_L_D2; 2.
DR Pfam; PF02787; CPase_L_D3; 1.
DR Pfam; PF02142; MGS; 1.
DR PRINTS; PR00098; CPaseII_lrg; 1.
DR TIGRfam; TIGR01369; CPaseII_lrg; 1.
DR PROSITE; PS00866; CPase_1; 2.
DR PROSITE; PS00867; CPase_2; 2.
DR Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
KW ATP-binding; Manganese; Complete proteome.
FT DOMAIN 1 401
FT DOMAIN 402 546
FT DOMAIN 547 929
FT DOMAIN 930 1057
FT REPEAT 1 546
FT REPEAT 547 1057
FT NP_BIND 153 210
FT NP_BIND 302 352
FT METAL 284 298
FT METAL 298 300
FT METAL 300 320
FT METAL 820 832
FT METAL 832 832
SQ SEQUENCE 1057 AA; 117171 MW; E3E179BF0591F0F8 CRC64;

Query Match 66.1%; Score 37; DB 1; Length 1057;
Best Local Similarity 60.0%; Pred. No. 16;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPGMHYS 11
DB 190 EIVSNGHYS 199

RESULT 4
CARB_STRAW STANDARD; PRT; 1057 AA.
AC P58940;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
GN phosphate synthetase ammonia chain).
OS Staphylococcus aureus (strain MW2).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=196620;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22040717; PubMed=12044378;
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
RA Nagai Y., Iwana N., Asano K., Naimi T., Kuroda H., Cui L.,
RA Yamamoto K., Hiramoto K.;
RT "Genome and virulence determinants of high virulence community-
RT acquired MRSA."
RL Lancet 359:1819-1827(2002).
CC -1- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
CC phosphate + L-glutamate + carbamoyl phosphate.
CC -1- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -1- PATHWAY: Arginine biosynthesis.
CC -1- PATHWAY: Pyrimidine biosynthesis.
CC -1- SUBUNIT: Composed of two chains; the small (or glutamine) chain
CC promotes the hydrolysis of glutamine to ammonia, which is used by
CC the large (or ammonia) chain to synthesize carbamoyl phosphate (By
CC similarity).
CC -1- SIMILARITY: BELONGS TO THE CARB FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

```

```

CC EMBL; AP004825; BAB94951.1; -.
DR HAMAP; MF 01210; -.
DR InterPro; IPR006275; CarA_L_glu.
DR InterPro; IPR005483; CPase_L.
DR InterPro; IPR005479; CPase_L_D2.
DR InterPro; IPR005480; CPase_L_D3.
DR InterPro; IPR005481; CPase_L_N.
DR InterPro; IPR004362; MGS like.
DR Pfam; PF00289; CPase_L_Chain; 2.
DR Pfam; PF02786; CPase_L_D2; 2.
DR Pfam; PF02787; CPase_L_D3; 1.
DR Pfam; PF02142; MGS; 1.
DR PRINTS; PR00098; CPase.
DR TIGRfam; TIGR01369; CPaseII_lrg; 1.
DR PROSITE; PS00866; CPase_1; 2.
DR PROSITE; PS00867; CPase_2; 2.
KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
KW ATP-binding; Manganese; Complete proteome.
FT DOMAIN 1 401
FT DOMAIN 402 546
FT DOMAIN 547 929
FT DOMAIN 930 1057
FT REPEAT 1 546
FT REPEAT 547 1057
FT NP_BIND 153 210
FT NP_BIND 302 352
FT METAL 284 298
FT METAL 298 300
FT METAL 300 320
FT METAL 820 820
FT METAL 832 832
SQ SEQUENCE 1057 AA; 117185 MW; D8E3B09F9BC6F152 CRC64;

Query Match 66.1%; Score 37; DB 1; Length 1057;
Best Local Similarity 60.0%; Pred. No. 16;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPGMHYS 11
DB 190 EIVSNGHYS 199

RESULT 5
Y11K_TYDVA STANDARD; PRT; 102 AA.
ID Y11K_TYDVA
AC P31619;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-OCT-1993 (Rel. 27, Last annotation update)
DE Hypothetical 11.2 kDa protein.
GN Y1.
OS Tobacco yellow dwarf virus (strain Australia) (TYDV).
OC Viruses; ssDNA viruses; Geminiviridae; Mastrevirus.
OX NCBI_TaxID=31599;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92185538; PubMed=1546458;
RA Morris B.A.M., Richardson K.A., Haley A., Zhan X., Thomas J.E.;
RT "The nucleotide sequence of the infectious cloned DNA component of
RT tobacco yellow dwarf virus reveals features of geminiviruses
RT infecting monocotyledonous plants."
RL Virology 187:633-642(1992).

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

```

```

DR PIR: A42452; A42452.
DR InterPro: IPR002621; Gemini mov.
DR Pfam: PF01708; Gemini_mov; I.
KW Hypothetical protein.
SQ SEQUENCE 102 AA; 11178 MW; A40ECF1E0AF55B67 CRC64;

Query Match 64.3%; Score 36; DB 1; Length 102;
Best Local Similarity 60.0%; Pred. No. 2.3; Mismatches 1; Indels 0; Gaps 0;
Matches 6; Conservative 3;

QY 2 EVVPGMGHYS 11
   :||| :|||
Db 7 QVPSGINS 16

RESULT 6
UMEL YEAST
ID UMEL YEAST STANDARD; PRT; 460 AA.
AC Q03010; P87330;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Meiosis negative regulator UME1.
GN UME1 OR WTM3 OR YPL139C OR LPI7C.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A364A;
RA Mallory M.J., Strich R.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RX MEDLINE=97313271; PubMed=9169875;
RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansoorge W.,
RA Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,
RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
RA Dietrich F.S., Delius H., DiPaolo T., Dubois E., Duesterhoeft A.,
RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
RA Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.,
RA Hunnicke-Smith S., Hyman R., Johnston M., Kalman S., Klein K.,
RA Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,
RA Marathe R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D.,
RA Mueller-Auer S., Namath A., Newtich U., Oefner P., Pearson D.,
RA Petel F.X., Pohl T.M., Purnelle D., Schafer M., Scharfe M.,
RA Scherens B., Schramm S., Schroeder M., Sidu A.M., Tettelin H.,
RA Urrestarazu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,
RA Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,
RA Zhong W.W., Zollner A., Vo D.H., Hani J.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";
RL Nature 387:103-105(1997).
CC -!- FUNCTION: TRANSCRIPTIONAL MODULATOR WITH ROLES IN MEIOTIC
CC REGULATION AND SILENCING. NEGATIVE REGULATOR OF MEIOSIS.
CC -!- SIMILARITY: Contains 4 WD repeats.
CC -!- SIMILARITY: STRONG, TO YEAST WTM1 AND WTM2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U10280; AAB40937.1; -
DR EMBL; U43703; AAB68221.1; -
DR PIR; S69046; S69046.
DR TRANSFAC; T04309; -
DR SGD; S0006060; UME1.

GO; GO:0005634; C:nucleus; IDA.
GO; GO:0003714; F:transcription co-repressor activity; IDA.
GO; GO:0040020; P:regulation of meiosis; IGI.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 3.
DR SMART; SM00320; WD40; 4.
DR PROSITE; PS00678; WD_REPEATS_1; FALSE NEG.
DR PROSITE; PS00083; WD_REPEATS_2; FALSE NEG.
DR PROSITE; PS00294; WD_REPEATS_REGION; FALSE NEG.
KW Transcription regulation; Meiosis; Repeat; WD repeat.
FT REPEAT 233 271 WD 1.
FT REPEAT 276 316 WD 2.
FT REPEAT 339 379 WD 3.
FT REPEAT 411 451 WD 4.
SQ SEQUENCE 460 AA; 51022 MW; AA6F60448B7BCBA9 CRC64;

Query Match 64.3%; Score 36; DB 1; Length 460;
Best Local Similarity 62.5%; Pred. No. 11;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMHY 10
   :||| :|||
Db 85 IVPLGLHY 92

RESULT 7
YK47 YEAST
ID YK47 YEAST STANDARD; PRT; 743 AA.
AC P36148;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 83.6 kDa protein in CCP1-MET1 intergenic region.
GN YKR067W.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c;
RA van Vliet-Reedijk J.C., Planta R.J.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -!- SIMILARITY: STRONG, TO YEAST YBL011W.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z28292; CAA82146.1; -
DR PIR; S38143; S38143.
DR SGD; S0001775; GFT2.
DR GO; GO:0005783; C:endoplasmic reticulum; IDA.
DR GO; GO:0004366; F:glycerol-3-phosphate O-acyltransferase acti. .; IDA.
DR GO; GO:0008654; P:phospholipid biosynthesis; IDA.
DR InterPro: IPR002123; Acyltransferase.
DR Pfam; PF01553; Acyltransferase; 1.
DR SMART; SM00563; Plsc; 1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 31 55 POTENTIAL.
FT TRANSMEM 69 85 POTENTIAL.
FT TRANSMEM 502 524 POTENTIAL.
FT TRANSMEM 539 555 POTENTIAL.
SQ SEQUENCE 743 AA; 83644 MW; 84B9946E56B82F15 CRC64;

Query Match 64.3%; Score 36; DB 1; Length 743;
Best Local Similarity 75.0%; Pred. No. 18;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

QY 3 VVPXGMHY 10
 ||| :||
 Db 294 VVPCGLHY 301

RESULT 8
 ID11 MESAU STANDARD; PRT; 227 AA.
 AC 035586;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Isopentenyl-diphosphate delta-isomerase 1 (EC 5.3.3.2) (IPP isomerase
 DE 1) (Isopentenyl pyrophosphate isomerase 1) (IPPI1).
 GN ID11.
 OS Mesocricetus auratus (Golden hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Mesocricetus.
 OX NCBI_TaxID=10036;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97373600; PubMed=9228075;
 RA Paton V.G., Shackelford J.E., Kriesans S.K.;
 RA "Cloning and subcellular localization of hamster and rat isopentenyl
 RT diphosphate dimethylallyl diphosphate isomerase. A Pts1 motif targets
 RT the enzyme to peroxisomes.";
 RL J. Biol. Chem. 272:18945-18950 (1997).
 CC -1- FUNCTION: CATALYZES THE 1,3-ALLYLIC REARRANGEMENT OF THE
 CC HOMOALLYLIC SUBSTRATE ISOPENTENYL (IPP) TO ITS HIGHLY
 CC ELECTROPHILIC ALLYLIC ISOMER, DIMETHYLLALLYL DIPHOSPHATE (DMAPP).
 CC -1- CATALYTIC ACTIVITY: Isopentenyl diphosphate = dimethylallyl
 CC diphosphate.
 CC -1- COFACTOR: REQUIRES MAGNESIUM FOR ACTIVITY.
 CC -1- PATHWAY: ISOPRENOID BIOSYNTHETIC PATHWAY WHOSE END PRODUCTS
 CC INCLUDE DOLICOLS, VITAMINS A, D, E, AND K, STEROID HORMONES,
 CC CAROTENOID BILE ACIDS AND CHOLESTEROL.
 CC -1- SUBCELLULAR LOCATION: Peroxisomal.
 CC -1- SIMILARITY: Belongs to the IPP isomerase type 1 family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AF003836; AAC53283.1; -;
 DR InterPro; IPR002667; IPP isomerase.
 DR Pfam; PF00293; NUDIX; 1.
 DR ProDom; PD004109; IPP isomerase; 1.
 KW Carotenoid biosynthesis; Cholesterol biosynthesis;
 KW Isoprene biosynthesis; Sterol biosynthesis; Isomerase; Peroxisome;
 KW Magnesium.
 FT ACT_SITE 86 86 BY SIMILARITY.
 FT ACT_SITE 148 148 BY SIMILARITY.
 FT SITE 225 227 MICROBODY TARGETING SIGNAL.
 SQ SEQUENCE 227 AA; 26317 MW; F500A6586385E803 CRC64;

Query Match 62.5%; Score 35; DB 1; Length 227;
 Best Local Similarity 70.0%; Pred. No. 8.4;
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMHY 10
 ||| :||
 Db 121 BEVDNEMHY 130

RESULT 9
 GAAB_METH STANDARD; PRT; 308 AA.
 AC 026806;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE GMP synthase [glutamine-hydrolyzing] subunit B (EC 6.3.5.2) (GMP
 DE synthetase).
 GN GAAB OR MTH710.
 OS Methanobacterium thermoautotrophicum.
 OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
 OC Methanobacteriaceae; Methanothermobacter.
 OX NCBI_TaxID=187420;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=Delta H;
 RA MEDLINE=98037514; PubMed=9371463;
 RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
 RA Alldredge T., Bahirzadeh R., Blakely D., Cook R., Gilbert K.,
 RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
 RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
 RA Jivani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
 RA McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
 RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
 RA "Complete genome sequence of Methanobacterium thermoautotrophicum
 RT deltaH: functional analysis and comparative genomics.";
 RL J. Bacteriol. 179:7135-7155 (1997).
 CC -1- CATALYTIC ACTIVITY: ATP + xanthosine 5'-phosphate + L-glutamine +
 CC H2O = AMP + diphosphate + GMP + L-glutamate.
 CC -1- PATHWAY: GMP biosynthesis.
 CC -1- SUBUNIT: HETERODIMER COMPOSED OF A GLUTAMINE AMIDOTRANSFERASE
 CC SUBUNIT (A) AND A GMP SYNTHASE SUBUNIT (B) (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE GMP SYNTHASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AE000850; AAB85215.1; ALT_INIT.
 DR HSRP; P04079; 1GPM.
 DR HAMAP; MF_00345; -; 1.
 DR InterPro; IPR001674; GMP synt_C.
 DR Pfam; PF00958; GMP synt_C; 1.
 DR TIGRFAMs; TIGR00884; guaA_Cterm; 1.
 KW Ligase; GMP biosynthesis; Purine biosynthesis; ATP-binding;
 KW Complete proteome.
 FT DOMAIN 33 184 GMP-BINDING (BY SIMILARITY).
 FT NP_BIND 29 35 ATP (BY SIMILARITY).
 SQ SEQUENCE 308 AA; 34403 MW; F2DC6FED202CABC1 CRC64;

Query Match 62.5%; Score 35; DB 1; Length 308;
 Best Local Similarity 63.6%; Pred. No. 12;
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMHY 11
 ||| :||
 Db 216 BEVVEGLHES 226

RESULT 10
 SLSI_YARLI STANDARD; PRT; 426 AA.
 AC 099156;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE SLSI protein precursor.
 GN SLSI.
 OS Yarrowia lipolytica (Candida lipolytica).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Dipodascaceae; Yarrowia.
 OX NCBI_TaxID=4952;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 20460 / W29;
 RX MEDLINE=96216076; PubMed=8662639;
 RA Boland A., Beckerich J.-M., Gaillardin C.;
 RT "Sislp, an endoplasmic reticulum component, is involved in the
 protein translocation process in the yeast *Yarrowia lipolytica*.";
 RL J. Biol. Chem. 271:11668-11675(1996).
 CC -!- FUNCTION: INVOLVED IN THE PROTEIN TRANSLATION PROCESS. MAY
 INTERACT DIRECTLY WITH TRANSLATING POLYPEPTIDES TO FACILITATE
 THEIR TRANSFER AND/OR HELP THEIR FOLDING IN THE ER. IT IS NOT
 REQUIRED FOR VIABILITY BUT IS ESSENTIAL FOR OPTIMAL GROWTH AT
 ELEVATED TEMPERATURES.
 CC -!- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL; Z50154; CAA90516.1; -;
 DR F01; S58132; S58132.
 DR InterPro; IPR000886; ER target.
 DR PROSITE; PS00014; ER TARGET; 1.
 KW Endoplasmic reticulum; Signal.
 FT SIGNAL 1 17 POTENTIAL.
 FT CHAIN 18 426 SISLP PROTEIN.
 FT SITE 423 426 PRESENT SECRETION FROM ER (POTENTIAL).
 SQ SEQUENCE 426 AA; 47201 MW; 0ACD7BF17540B8E2 CRC64;
 Query Match 60.7%; Score 34; DB 1; Length 426;
 Best Local Similarity 44.4%; Pred. No. 26;
 Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
 QY 1 BEVVPXGMH 9
 DB 52 DQVIPAGLH 60
 RESULT 11
 ID YL15_VIBPA STANDARD; PRT; 441 AA.
 AC P46231;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Hypothetical protein VP2115 (ORF3).
 GN VP2115.
 OS Vibrio parahaemolyticus.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 OX NCBI_TaxID=670;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RIMD 2210633 / Serotype O3:K6;
 RX MEDLINE=22508454; PubMed=13620739;
 RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
 RA Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
 RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
 RT "Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism
 distinct from that of *V. cholerae*.";
 RL Lancet 361:743-749(2003).
 RN [2]
 RP SEQUENCE OF 1-140 FROM N.A.
 RC STRAIN=BB22;
 RX MEDLINE=94292449; PubMed=8021208;
 RA McCarter L.L.;
 RT "MorY, a component of the sodium-type flagellar motor.";

J. Bacteriol. 176:4219-4225(1994).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -!- SIMILARITY: STRONG, TO H.INFLUENZAE H10325.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AP005080; BAC60378.1; -;
 DR EMBL; U06949; AAA21571.1; -;
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 21 41 POTENTIAL.
 FT TRANSMEM 51 71 POTENTIAL.
 FT TRANSMEM 94 114 POTENTIAL.
 FT TRANSMEM 118 138 POTENTIAL.
 FT TRANSMEM 150 170 POTENTIAL.
 FT TRANSMEM 195 215 POTENTIAL.
 FT TRANSMEM 239 259 POTENTIAL.
 FT TRANSMEM 260 280 POTENTIAL.
 FT TRANSMEM 291 311 POTENTIAL.
 FT TRANSMEM 334 354 POTENTIAL.
 FT TRANSMEM 363 383 POTENTIAL.
 FT TRANSMEM 419 439 POTENTIAL.
 SQ SEQUENCE 441 AA; 45961 MW; 451969FE307E4D46 CRC64;
 Query Match 58.9%; Score 33; DB 1; Length 441;
 Best Local Similarity 54.5%; Pred. No. 43;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 1 BEVVPXGMH 11
 DB 417 ETVPVTEHYN 427
 RESULT 12
 ID PHSL_DESBA STANDARD; PRT; 513 AA.
 AC P13065;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Periplasmic [Nifese] hydrogenase large subunit (EC 1.12.99.6) (Nifese
 hydrogenase large chain).
 OS Desulfovibrio baculatus (Desulfomicrobium baculatus).
 OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfovibrionales;
 OC Desulfomicrobiaceae; Desulfomicrobium.
 OX NCBI_TaxID=899;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88058744; PubMed=3316183;
 RA Menon N.K., Peck H.D. Jr., le Gall J., Przybyla A.E.;
 RT "Cloning and sequencing of the genes encoding the large and small
 subunits of the periplasmic (Nifese) hydrogenase of *Desulfovibrio*
baculatus.";
 RL J. Bacteriol. 169:5401-5407(1987).
 RN [2]
 RP REVISIONS.
 RA Menon N.K., Peck H.D. Jr., le Gall J., Przybyla A.E.;
 RL J. Bacteriol. 170:4429-4439(1988).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS).
 RX MEDLINE=99306038; PubMed=10378275;
 RA Garcin E., Verne X., Hatchikian E.C., Volbeda A., Frey M.,
 RA Pontecilla-Camps J.C.;
 RT "Removal of the bridging ligand atom at the Ni-Fe active site of
 [Nifese] hydrogenase upon reduction with H2, as revealed by X-ray
 structure analysis at 1.4-A resolution.";
 RL Structure 7:557-566(1999).
 CC -!- CATALYTIC ACTIVITY: H(2) + acceptor = 2 H(+) + reduced acceptor.

CC -1- COFACTOR: Nickel, 2 irons and selenocysteine. Iron 1 has three
CC cyanide and carbon monoxide ligands. Iron 2 has three water
CC ligands.
CC -1- SUBUNIT: Heterodimer of a large and a small subunit.
CC -1- SUBCELLULAR LOCATION: Periplasmic.
CC -1- MISCELLANEOUS: PERHAPS THE LEADER OF THE SMALL SUBUNIT SERVES AS A
CC TRANSPORT VEHICLE FOR BOTH SUBUNITS.
CC -1- SIMILARITY: BELONGS TO THE [NIFE] [NIFESE] HYDROGENASE LARGE
CC SUBUNIT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M18271; AAA23375.2; -;
CC PIR; A33101; HQDVLB.
CC DR PDB; 1CC1; 01-JUN-99.
CC DR InterPro; IPR001501; Ni_hdl.
CC DR Pfam; PF00374; Nifese_Hases; 1.
CC DR PROSITE; PS00507; NI_HGENASE L1; 1.
CC DR PROSITE; PS00508; NI_HGENASE L2; 1.
KW Oxidoreductase; Periplasmic; Metal-binding; Nickel; Iron; Selenium;
KW Selenocysteine; 3D-structure.
FT INIT_MET 0 0
FT METAL 51 51 IRON 2.
FT METAL 70 70 NICKEL.
FT METAL 73 73 IRON 1.
FT METAL 73 73 NICKEL.
FT METAL 444 444 IRON 2 (VIA CARBONYL OXYGEN).
FT METAL 492 492 NICKEL.
FT METAL 495 495 IRON 1.
FT METAL 498 498 NICKEL.
FT METAL 498 498 IRON 2.
FT SE_CYS 492 492
FT STRAND 13 16
FT STRAND 23 23
FT STRAND 26 33
FT TURN 34 35
FT STRAND 36 44
FT STRAND 46 46
FT HELIX 50 53
FT TURN 54 56
FT HELIX 59 61
FT HELIX 62 65
FT HELIX 66 69
FT HELIX 74 89
FT TURN 90 90
FT HELIX 95 119
FT TURN 120 121
FT HELIX 122 124
FT TURN 125 125
FT TURN 132 133
FT TURN 139 140
FT HELIX 144 147
FT HELIX 151 183
FT STRAND 192 193
FT TURN 194 195
FT STRAND 196 197
FT HELIX 203 222
FT TURN 223 223
FT HELIX 224 234
FT STRAND 236 239
FT STRAND 248 250
FT STRAND 253 255
FT TURN 258 259
FT STRAND 263 265
FT STRAND 268 271
FT TURN 272 273
FT STRAND 274 276

FT HELIX 280 282
FT STRAND 283 286
FT TURN 288 289
FT STRAND 290 290
FT TURN 291 291
FT STRAND 292 292
FT STRAND 300 301
FT HELIX 302 304
FT STRAND 309 309
FT TURN 311 312
FT TURN 314 315
FT STRAND 318 318
FT STRAND 322 325
FT TURN 326 327
FT STRAND 328 328
FT STRAND 331 331
FT HELIX 334 341
FT STRAND 344 344
FT HELIX 347 357
FT STRAND 363 363
FT HELIX 364 371
FT TURN 374 398
FT TURN 401 402
FT STRAND 406 406
FT STRAND 415 423
FT TURN 424 425
FT STRAND 426 435
FT TURN 436 437
FT STRAND 438 446
FT TURN 447 447
FT HELIX 448 451
FT TURN 457 458
FT HELIX 463 468
FT TURN 469 470
FT STRAND 472 472
FT TURN 476 477
FT HELIX 480 488
FT TURN 489 489
FT STRAND 492 492
FT HELIX 493 497
SQ SEQUENCE 513 AA; 56683 MW; AC8285A6F80576FC CRC64;

Query Match 58.9%; Score 33; DB 1; Length 513;
Best Local Similarity 71.4%; Pred. No. 50;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 5 PEXGMHYS 11
|:|:|
Db 297 PGGLHYS 303

RESULT 13
MUTL_BACSU STANDARD; PRT; 627 AA.
AC P49850;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA mismatch repair protein mutL.
GN MUTL.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=168;
RX MEDLINE=96349107; PubMed=8760914;
RA Ginetti F., Perego M., Albertini A.M., Galizzi A.;
RT "Bacillus subtilis mutS mutL operon: identification, nucleotide
sequence and mutagenesis.";
RL Microbiology 142:2021-2029 (1996).
RN [2]
RP SEQUENCE FROM N.A.


```

RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bortner M.G., Bessieres P., Bolotin A., Borchert S.,
RA Boreiss R., Bourlier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Chou S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Davine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Gallier N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Meene D., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone N., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Ruchelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche B., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terstra P., Tognoni A.,
RA Tostato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambut R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis"; 249-256(1997).
RL Nature 390:249-256(1997).
CC -!- FUNCTION: THIS PROTEIN IS INVOLVED IN THE REPAIR OF MISMATCHES IN
CC DNA. IT IS REQUIRED FOR DAM-DEPENDENT METHYL-DIRECTED DNA MISMATCH
CC REPAIR. MAY ACT AS A "MOLECULAR MATCHMAKER", A PROTEIN THAT
CC PROMOTES THE FORMATION OF A STABLE COMPLEX BETWEEN TWO OR MORE
CC DNA-BINDING PROTEINS IN AN ATP-DEPENDENT MANNER WITHOUT ITSELF
CC BEING PART OF A FINAL EFFECTOR COMPLEX (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MUTL/HEXB FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; U27343; AAB19236.1; -
DR EMBL; Z59112; CAB13578.1; -
DR PIR; A69663; A69663.
DR HSSP; P23367; 1BKN.
DR HAMAP; MF_00149; -; 1.
DR HAMAP; BG11402; mutL.
DR InterPro; IPR003594; ATPbind_ATPase.
DR InterPro; IPR002099; DNA_mis_repair.
DR Pfam; PF01119; DNA_mis_repair; 1.
DR Pfam; PF02518; HATPase_C; 1.
DR SMART; SM00387; HATPase_C; 1.
DR TIGRFAMs; TIGR00595; mutL; 1.
DR PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
DR DNA repair; Complete proteome.
KW SEQUENCE 627 AA; 70431 MW; 068A0509C265343 CRC64;

Query Match 58.9%; Score 33; DB 1; Length 627;
Best Local Similarity 54.5%; Pred. No. 62;
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 REVVPXGMHYS 11
DB 488 EMIVPLTHYS 498

-----
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bortner M.G., Bessieres P., Bolotin A., Borchert S.,
RA Boreiss R., Bourlier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Chou S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Davine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Gallier N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Meene D., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone N., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Ruchelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche B., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terstra P., Tognoni A.,
RA Tostato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambut R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis"; 249-256(1997).
RL Nature 390:249-256(1997).
CC -!- FUNCTION: THIS PROTEIN IS INVOLVED IN THE REPAIR OF MISMATCHES IN
CC DNA. IT IS REQUIRED FOR DAM-DEPENDENT METHYL-DIRECTED DNA MISMATCH
CC REPAIR. MAY ACT AS A "MOLECULAR MATCHMAKER", A PROTEIN THAT
CC PROMOTES THE FORMATION OF A STABLE COMPLEX BETWEEN TWO OR MORE
CC DNA-BINDING PROTEINS IN AN ATP-DEPENDENT MANNER WITHOUT ITSELF
CC BEING PART OF A FINAL EFFECTOR COMPLEX (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MUTL/HEXB FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; U27343; AAB19236.1; -
DR EMBL; Z59112; CAB13578.1; -
DR PIR; A69663; A69663.
DR HSSP; P23367; 1BKN.
DR HAMAP; MF_00149; -; 1.
DR HAMAP; BG11402; mutL.
DR InterPro; IPR003594; ATPbind_ATPase.
DR InterPro; IPR002099; DNA_mis_repair.
DR Pfam; PF01119; DNA_mis_repair; 1.
DR Pfam; PF02518; HATPase_C; 1.
DR SMART; SM00387; HATPase_C; 1.
DR TIGRFAMs; TIGR00595; mutL; 1.
DR PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
DR DNA repair; Complete proteome.
KW SEQUENCE 627 AA; 70431 MW; 068A0509C265343 CRC64;

Query Match 58.9%; Score 33; DB 1; Length 627;
Best Local Similarity 54.5%; Pred. No. 62;
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 REVVPXGMHYS 11
DB 488 EMIVPLTHYS 498

```

```

RESULT 14
KPBA_CABEL STANDARD; PRT; 1188 AA.
ID KP34335;
AC 01-FEB-1994 (Rel. 28, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable phosphorylase B kinase alpha regulatory chain (Phosphorylase
DE Kinase alpha subunit).
DE GN C14B9.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
[1]
SEQUENCE FROM N.A.
RP STRAIN=Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laisster N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy N., Rifkin L., Roopra A., Saunders D., Shownkeen R.,
RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP REVISIONS.
RA Waterston R.;
RA Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PHOSPHORYLASE B KINASE CATALYZES THE PHOSPHORYLATION OF
CC SERINE IN CERTAIN SUBSTRATES INCLUDING TROPONIN I. THE ALPHA
CC CHAIN MAY BIND CALMODULIN (BY SIMILARITY).
CC -!- PATHWAY: Glycogen metabolism.
CC -!- SIMILARITY: BELONGS TO THE PHOSPHORYLASE B KINASE REGULATORY
CC CHAINS FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; L15188; AAA27954.2; -
DR WormPep; C14B9.8; CE26870.
KW Hypothetical protein; Glycogen metabolism; Calmodulin-binding.
SQ SEQUENCE 1188 AA; 135558 MW; DE9BB875F3603863 CRC64;

Query Match 58.9%; Score 33; DB 1; Length 1188;
Best Local Similarity 54.5%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 REVVPXGMHYS 11
DB 950 BEYADGIHYS 960

-----
RESULT 15
VLTP_BPT5 STANDARD; PRT; 1396 AA.
ID VLTP_BPT5
AC P13350; O48502;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)

```



```

DE L-shaped tail fiber protein (LTF protein).
GN LTF.
OS Bacteriophage T5.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
OC T5-like viruses.
OX NCBI_TaxID=10726;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95309401; PubMed=7789514;
RA Kallman A.V., Kulshin V.B., Shlyapnikov M.G., Keenzenko V.N.,
RA Kryukov V.M.;
RT "The nucleotide sequence of the bacteriophage T5 ltf gene.";
RL FEBS Lett. 366:46-48(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA Kallman A.V.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=88289370; PubMed=3267228;
RA Kallman A.V., Kryukov V.M., Bayev A.A.;
RT "The nucleotide sequence of bacteriophage T5 DNA at the region
RT between early and late genes.";
RL Nucleic Acids Res. 16:6230-6230(1988).
CC -!- FUNCTION: NONESSENTIAL PROTEIN THAT MEDIATES BINDING TO THE
CC POLYMANNOSE O ANTIGEN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X69460; CAA49220.1; -.
DR EMBL; AJ001191; CAA04591.1; -.
DR PIR; S65934; S36851.
KW Late protein.
FT CONFLICT 986 986 V -> A (IN REF. 2).
SQ SEQUENCE 1396 AA; 147989 MW; 18CD2192F65FFFC1 CRC64;

```

```

Query Match      58.9%; Score 33; DB 1; Length 1396;
Best Local Similarity 50.0%; Pred. No. 1.4e+02;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

```

```

Oy      2 EVVPXGMHYS 11
       : : |||||
Db      1360 KTIPAGDHYS 1369

```

Search completed: December 22, 2003, 17:42:26
Job time : 4.6 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 17:27:26 ; Search time 25.2 Seconds
(without alignments)
112.642 Million cell updates/sec

Title: US-09-909-164-11
Perfect score: 56
Sequence: 1 BEVVPXGMHYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | ID | Description |
|------------|-------|-------|--------|-----------|---------------------|
| 1 | 41 | 73.2 | 413 | 11 Q8K289 | Q8K289 mus musculus |
| 2 | 39 | 69.6 | 1057 | 16 Q8CPJ4 | Q8CPJ4 staphylococ |
| 3 | 37 | 66.1 | 308 | 16 Q9X2E2 | Q9X2E2 thermotoga |
| 4 | 37 | 66.1 | 322 | 17 Q9HLH8 | Q9HLH8 thermoplas |
| 5 | 37 | 66.1 | 1044 | 16 Q8DIH0 | Q8DIH0 synechococ |
| 6 | 36 | 64.3 | 208 | 2 Q46486 | Q46486 corynebacte |
| 7 | 36 | 64.3 | 252 | 16 Q8YWP1 | Q8YWP1 anabaena sp |
| 8 | 36 | 64.3 | 819 | 10 Q9AVK4 | Q9AVK4 pisum sativ |
| 9 | 35 | 62.5 | 139 | 2 Q57489 | Q57489 bacteroides |
| 10 | 35 | 62.5 | 156 | 16 Q9PC35 | Q9PC35 xyella fas |
| 11 | 35 | 62.5 | 233 | 10 Q40479 | Q40479 nicotiana t |
| 12 | 35 | 62.5 | 237 | 10 Q3LW50 | Q3LW50 nicotiana s |
| 13 | 35 | 62.5 | 317 | 9 Q38317 | Q38317 lactobacill |
| 14 | 35 | 62.5 | 425 | 5 Q9XVK4 | Q9XVK4 caenorhabdi |
| 15 | 35 | 62.5 | 510 | 10 Q9SA71 | Q9SA71 arabidopsis |
| 16 | 35 | 62.5 | 2042 | 17 Q8TZ07 | Q8TZ07 methanopyru |

17 34 60.7 264 17 027902
18 34 60.7 279 16 09RXN9
19 34 60.7 285 16 098HU6
20 34 60.7 350 16 09RW92
21 34 60.7 355 16 09XAM3
22 34 60.7 360 17 027679
23 34 60.7 425 16 09KBA1
24 34 60.7 484 11 08CAL1
25 34 60.7 495 5 016912
26 34 60.7 637 5 03Y0Y6
27 34 60.7 678 12 09BIX6
28 34 60.7 748 12 09YR01
29 34 60.7 992 11 08BNL0
30 34 60.7 1028 16 08YJ11
31 34 60.7 1070 11 08C6X2
32 34 60.7 2438 5 09VQL7
33 33 58.9 156 3 012479
34 33 58.9 158 16 08P6P8
35 33 58.9 208 2 08KTQ4
36 33 58.9 216 16 09RDC1
37 33 58.9 254 16 08PIB0
38 33 58.9 257 8 099011
39 33 58.9 273 4 096N44
40 33 58.9 306 16 08XVQ7
41 33 58.9 364 5 08IGN5
42 33 58.9 367 16 0916J1
43 33 58.9 387 16 098FX1
44 33 58.9 425 16 08D881
45 33 58.9 426 3 09HG99

ALIGNMENTS

RESULT 1

Q8K289 PRELIMINARY; PRT; 413 AA.
AC Q8K289;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to expressed sequence A1987856.
GN 1110004B19RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strauberg R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBSJ databases.
DR EMBL, BC032195; AAH32195.1; -;
DR MGD; MGI:1915724; 1110004B19RIK.
SQ SEQUENCE 413 AA; 46621 MW; A9B8A1DC70CDA0D5 CRC64;

Query Match 73.2%; Score 41; DB 11; Length 413;
Best Local Similarity 63.6%; Pred. No. 5.4;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVVPXGMHYS 11
||: |||||
Db 168 BEIPPLGKHYS 178

RESULT 2

Q8CPJ4 PRELIMINARY; PRT; 1057 AA.
ID Q8CPJ4;
AC Q8CPJ4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Carbamoyl-phosphate synthase large chain.
GN SE08079.

OS Staphylococcus epidermidis.
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 OX NCBI_TaxID=1282;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 12228;
 RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
 RA Chen Z., Wen Y.;
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE016746; AAO04476.1; -
 KW Complete proteome.
 SQ SEQUENCE 1057 AA; 117391 MW; 8944D7D8DBICAB59 CRC64;

Query Match 69.6%; Score 39; DB 16; Length 1057;
 Best Local Similarity 63.6%; Pred. No. 38;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 1 EEVVPXGMHY 11
 :||| :|||
 Db 189 KEVVSNGLHY 199

RESULT 3
 Q9X2E2 PRELIMINARY; PRT; 308 AA.
 AC Q9X2E2;
 DT 01-NOV-1999 (TRENBLrel. 12, Created)
 DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE FTS protease activity modulator HFLK.
 GN TM1822.
 OS Thermotoga maritima.
 OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
 OX NCBI_TaxID=2336;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MSB8 / DSM 3109;
 RX MEDLINE=99287316; PubMed=10360571;
 RA Nelson K.E., Clayton R.A., Gill S.R., Winn M.L., Dodson R.J.,
 RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
 RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
 RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Bisen J.A., White O.,
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from
 RT genome sequence of Thermotoga maritima."
 RL Nature 399:323-329 (1999).
 DR EMBL; AE001819; AAD36885.1; -
 DR TIGR; TM1822; -
 DR InterPro; IPR001107; Band 7.
 DR InterPro; IPR001972; Stomatatin.
 DR Pfam; PF01145; Band 7; 1.
 DR PRINTS; PR00721; STOMATIN.
 DR SMART; SM00244; PHB; 1.
 KW Protease; Complete proteome.
 SQ SEQUENCE 308 AA; 34778 MW; ADE03603E5101A9D CRC64;

Query Match 66.1%; Score 37; DB 16; Length 308;
 Best Local Similarity 75.0%; Pred. No. 26;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMHY 10
 :||| :|||
 Db 41 VVPSGIHY 48

RESULT 4
 Q9HLH8 PRELIMINARY; PRT; 322 AA.
 ID Q9HLH8
 AC Q9HLH8;
 DT 01-MAR-2001 (TRENBLrel. 16, Created)
 DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)

DE Glucose-fructose oxidoreductase related protein.
 GN TA0250.
 OS Thermoplasma acidophilum.
 OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
 OC Thermoplasmataceae; Thermoplasma.
 OX NCBI_TaxID=2303;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DSM 1728;
 RX MEDLINE=20479972; PubMed=11029001;
 RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
 RA Meves H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
 RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
 RT acidophilum."
 RL Nature 407:508-513 (2000).
 DR EMBL; AL445063; CAC11395.1; -
 DR InterPro; IPR000683; GFO_IDH_MoCA.
 DR Pfam; PF01408; GFO_IDH_MoCA; 1.
 KW Complete proteome.
 SQ SEQUENCE 322 AA; 36918 MW; B8C239E71009D167 CRC64;

Query Match 66.1%; Score 37; DB 17; Length 322;
 Best Local Similarity 75.0%; Pred. No. 27;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMHY 10
 :||| :|||
 Db 66 VVPGLHY 73

RESULT 5
 Q8DIH0 PRELIMINARY; PRT; 1044 AA.
 ID Q8DIH0
 AC Q8DIH0;
 DT 01-MAR-2003 (TRENBLrel. 23, Created)
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE Multidrug efflux transporter.
 GN TL1618.
 OS Synechococcus elongatus (Thermosynechococcus elongatus).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
 OX NCBI_TaxID=32046;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BP-1;
 RX MEDLINE=2225144; PubMed=12240834;
 RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
 RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
 RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
 RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the thermophilic cyanobacterium
 RT Thermosynechococcus elongatus BP-1."
 RL DNA Res. 9:123-130 (2002).
 DR EMBL; AP005374; BAC09170.1; -
 KW Complete proteome.
 SQ SEQUENCE 1044 AA; 113205 MW; 00B9C13F0F636D2F CRC64;

Query Match 66.1%; Score 37; DB 16; Length 1044;
 Best Local Similarity 63.6%; Pred. No. 98;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGMHY 11
 :||| :|||
 Db 843 EEVLPNGIGYS 853

RESULT 6
 Q46486 PRELIMINARY; PRT; 208 AA.
 ID Q46486
 AC Q46486;
 DT 01-NOV-1996 (TRENBLrel. 01, Created)
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
 DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)

DE Hypothetical 23.0 kDa protein (GcrA).
GN GCR.
OS Corynebacterium xerosis, and
OS Corynebacterium striatum.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=1725, 43770;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=C.xerosis; STRAIN=M82B;
RY MEDLINE=96117603; PubMed=8559800;
RA Tauch A., Kassing F., Kalinowski J., Puhler A.;
RT "The Corynebacterium xerosis composite transposon Tn5432 consists of
RT two identical insertion sequences, designated IS1249, flanking the
RT erythromycin resistance gene ermC.";
RL Plasmid 34:119-131(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=C.striatum; STRAIN=M82B;
RX MEDLINE=20194806; PubMed=10732668;
RA Tauch A., Krieff S., Kalinowski J., Puhler A.;
RT "The 51,409-bp R-plasmid pRP10 from the multiresistant clinical
RT isolate Corynebacterium striatum M82B is composed of DNA segments
RT initially identified in soil bacteria and in plant, animal, and human
RT pathogens.";
RL Mol. Gen. Genet. 263:11-11(2000).
DR EMBL; U21300; AAC95478.1; -.
DR EMBL; AF024666; AAG03390.1; -.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 208 AA; 23012 MW; F1504BEECD895A6 CRC64;
Query Match 64.3%; Score 36; DB 2; Length 208;
Best Local Similarity 50.0%; Pred. No. 27;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
QY 2 EVVPXGMHYS 11
:|:|:|:|:
DB 130 DVIPEKHYA 139
RESULT 7
QBYWPI PRELIMINARY; PRT; 252 AA.
ID Q8YWP1
AC Q8YWP1; 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DE Hypothetical protein Alr1563.
GN ALR1563.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Kawanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL; AF003586; BAB77929.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 252 AA; 28831 MW; 925572DA5D1CA519 CRC64;
Query Match 64.3%; Score 36; DB 16; Length 252;
Best Local Similarity 50.0%; Pred. No. 34;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
QY 1 BEVVPXGMHY 10
:|:|:|:|:
DB 235 EMIVPAGLHF 244
RESULT 8
Q9AVK4 PRELIMINARY; PRT; 819 AA.
ID Q9AVK4
AC Q9AVK4; 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE SCARECROW.
GN PSRCR.
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids 1; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.
OX NCBI_TaxID=3888;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Alaska;
RX MEDLINE=21231727; PubMed=11333309;
RA Sassa N., Matsushita Y., Nakamura T., Nyunoya H.;
RT "The Molecular Characterization and in situ Expression Pattern of Pea
RT SCARECROW Gene.";
RL Plant Cell Physiol. 42:385-394(2001).
DR EMBL; AB048713; BAB39155.1; -.
DR InterPro; IPR001444; Flag_bb_rod.
DR InterPro; IPR005202; GRAS.
DR Pfam; PF03514; GRAS; 1.
DR PROSITE; PS00588; FLAGELLA_BB_ROD; 1.
SQ SEQUENCE 819 AA; 90372 MW; 41B67BD6DC72ADFA CRC64;
Query Match 64.3%; Score 36; DB 10; Length 819;
Best Local Similarity 45.5%; Pred. No. 1.2e+02;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
QY 1 BEVVPXGMHYS 11
:|:|:|:|:
DB 343 DDVVPTSLHFS 353
RESULT 9
Q57489 PRELIMINARY; PRT; 139 AA.
ID Q57489
AC Q57489; 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE DNA ligase (Fragment).
OS Bacteroides nodosus (Dichelobacter nodosus).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Cardiobacteriales;
OC Cardiobacteriaceae; Dichelobacter.
OX NCBI_TaxID=870;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96020672; PubMed=7476204;
RA Moses E.K., Good R.T., Sinistaj M., Billington S.J., Langford C.J.,
RA Rood J.I.;
RT "A multiple site-specific DNA-inversion model for the control of Omp1
RT phase and antigenic variation in Dichelobacter nodosus.";
RL Mol. Microbiol. 17:183-196(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96257263; PubMed=8654969;
RA Billington S.J., Sinistaj M., Cheetham B.F., Ayres A., Moses E.K.,
RA Katz M.E., Rood J.I.;
RT "Identification of a native Dichelobacter nodosus plasmid and
RT implications for the evolution of the vap regions.";
RL Gene 172:111-116(1996).
DR EMBL; U02462; AAB12366.1; -.
DR InterPro; IPR001357; BRCT.
DR Pfam; PF00533; BRCT; 1.
DR SMART; SM00292; BRCT; 1.

DR PROSITE; PS50172; BRCT; 1.
 FT NON-TER 1
 SQ SEQUENCE 139 AA; 15052 MW; E0E110AA4B7D4708 CRC64;

Query Match 62.5%; Score 35; DB 2; Length 139;
 Best Local Similarity 55.6%; Pred. No. 28;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMHYS 11
 Db 21 IVPAGVHWS 29

RESULT 10

ID Q9PC35 PRELIMINARY; PRT; 156 AA.
 AC Q9PC35;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical protein Xf1950.
 GN Xf1950.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xylella.
 OX NCBI_TaxID=2371;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9a5c;
 RX MEDLINE=20365717; PubMed=10910347;
 RA Simpson A.J.G., Reinach P.C., Arruda P., Abreu P.A., Acencio M.,
 RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrier H.,
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lenos E.G.M., Lenos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marinho C.L.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pasquero J.B.,
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 RA da Silva J.F., Truzzi M.F., Truzzi D., Tsai S.M., Tsuchioka M.H.,
 RA de Souza A.P., Terenzi M.F., Truzzi D., Tsai S.M., Tsuchioka M.H.,
 RA Vallada H., Van Sluys M.A., Verjovsky-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
 RT "The genome sequence of the plant pathogen Xylella fastidiosa";
 RL Nature 406:151-159(2000).
 DR EMBL; AB004014; AAF84752.1; -;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 156 AA; 17144 MW; D8358619C6671A5D CRC64;

Query Match 62.5%; Score 35; DB 16; Length 156;
 Best Local Similarity 55.6%; Pred. No. 32;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 BEVVPXGMH 9
 Db 119 BEILPQGVH 127

RESULT 11

Qy 1 BEVVPXGMH 9
 Db 119 BEILPQGVH 127

Q04079

ID Q04079 PRELIMINARY; PRT; 233 AA.
 AC Q04079;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE ERBBP-2.
 OS Nicotiana tabacum (Common tobacco).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamids; Solanales; Solanaceae; Nicotiana.
 OX NCBI_TaxID=4097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BY4; TISSUE=Leaf;
 RX MEDLINE=95276459; PubMed=7756828;
 RA Ohme-Takagi M., Shinshi H.;
 RT "Ethylene-inducible DNA binding proteins that interact with an
 RT ethylene responsive element";
 RL Plant Cell 7:173-182(1995).
 DR EMBL; D38126; BAA07324.1; -;
 DR HSSP; O80337; ZGCC.
 DR TRANSFAC; T02654; -;
 DR InterPro; IPR001471; TF_ERF.
 DR Pfam; PF00847; AP2-domain; 1.
 DR ProDom; PD001423; TF_ERF; 1.
 DR SMART; SM00380; AP2_1.
 SQ SEQUENCE 233 AA; 25563 MW; 6CD16783582C0CB5 CRC64;

Query Match 62.5%; Score 35; DB 10; Length 233;
 Best Local Similarity 60.0%; Pred. No. 50;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVVPXGMH 10

Db 90 QAVVPKGRHY 99

RESULT 12

ID Q9LW50 PRELIMINARY; PRT; 237 AA.
 AC Q9LW50;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Ethylene-responsive element binding factor.
 GN NSRFP2.
 OS Nicotiana sylvestris (Wood tobacco).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamids; Solanales; Solanaceae; Nicotiana.
 OX NCBI_TaxID=4096;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20399450; PubMed=10945353;
 RA Kitajima S., Koyama T., Ohme-Takagi M., Shinshi H., Sato F.;
 RT "Characterization of gene expression of NSRFPs, transcription factors
 RT of basic PR genes from Nicotiana sylvestris";
 RL Plant Cell Physiol. 41:817-824(2000).
 DR EMBL; AB016264; BAA97122.1; -;
 DR HSSP; O80337; ZGCC.
 DR InterPro; IPR001471; TF_ERF.
 DR Pfam; PF00847; AP2-domain; 1.
 DR ProDom; PD001423; TF_ERF; 1.
 DR SMART; SM00380; AP2_1.
 SQ SEQUENCE 237 AA; 26243 MW; 01BC3EEB51B46298 CRC64;

Query Match 62.5%; Score 35; DB 10; Length 237;
 Best Local Similarity 60.0%; Pred. No. 51;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVVPXGMH 10

Db 90 QAVVPKGRHY 99

Db 94 QAVPKGRHY 103

RESULT 13

Q38317 ID Q38317 PRELIMINARY; PRT; 317 AA.
 AC Q38317
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Lysin.
 GN Lys.
 OS Lactobacillus bacteriophage phi adh.
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
 OX NCBI_TaxID=12417;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Altermann E.;
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99384014; PubMed=10452953;
 RA Altermann E., Klein J., Henrich B.;
 RT "Primary structure and features of the genome of the Lactobacillus
 RT gasseri temperate bacteriophage phi-adh.";
 RL Gene 236:333-346(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99138034; PubMed=7836307;
 RA Henrich B., Binshofer B., Blaesi U.;
 RT "Primary structure and functional analysis of the lysis genes of
 RT Lactobacillus gasseri bacteriophage phi-adh.";
 RL J. Bacteriol. 177:723-732(1995).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93231538; PubMed=8472961;
 RA Fremaux C., De Antoni G., Raya R., Klaenhammer T.;
 RT "Genetic organization and sequence of the region encoding integrative
 RT functions from Lactobacillus gasseri temperate bacteriophage phi-
 RT adh.";
 RL Gene 126:61-66(1993).
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Engel G., Altermann E., Klein J., Henrich B.;
 RT "Structure of a genome region of the Lactobacillus gasseri temperate
 RT phage phi adh covering a repressor gene and cognate promoters.";
 RL Gene 251:67-70(1998).
 DR EMBL; AJ131519; CAB52540.1; -;
 DR InterPro; IPR002053; Glyco_hydro_25.
 DR Pfam; PF01183; Glyco_hydro_25; 1.
 DR ProDom; PD004620; Glyco_hydro_25; 1.
 DR SMART; SM00641; Glyco_25; 1.
 DR SMART; SM00287; SH3b; 1.
 SQ SEQUENCE 317 AA; 34703 MW; 9FP2715EB43561C7 CRC64;

Query Match 62.5%; Score 35; DB 9; Length 317;
 Best Local Similarity 66.7%; Pred. No. 69;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 3 VVPXGMHYS 11
 |||||
 Db 60 VVPMGYHYA 68

RESULT 14

Q9XVK4 ID Q9XVK4 PRELIMINARY; PRT; 425 AA.
 AC Q9XVK4
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE R10D12.10 protein.

GN R10D12.10.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Percy C.M.;
 RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C. elegans: A platform for
 RT investigating biology.";
 RL Science 282:2012-2018(1998).
 DR EMBL; Z81109; CAB03241.1; -;
 DR WormPep; R10D12.10; CE12690.
 DR InterPro; IPR000719; Prot_kinase.
 DR Pfam; PF00069; pkinase; 1.
 DR ProDom; PD000001; Prot_kinase; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 KW ATP-binding; Transferase.
 SQ SEQUENCE 425 AA; 49410 MW; 5D96E29B08C9E9D6 CRC64;

Query Match 62.5%; Score 35; DB 5; Length 425;
 Best Local Similarity 50.0%; Pred. No. 95;
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 BEVVPXGMHY 10
 |||||
 Db 335 EQIVPGGLQY 344

RESULT 15

Q9SA71 ID Q9SA71 PRELIMINARY; PRT; 510 AA.
 AC Q9SA71;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE T518.1 protein.
 GN T518.1.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,
 Li J., Kremenetskaia I., Luros J., Ngan I., Gonzalez A., Altati H.,
 Araujo R., Chao Q., Conn L., Conway A.B., Dunn P., Hansen N.,
 Huizar L., Kim C., Palm C., Rowley D., Shinn P., Walker M.,
 Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;
 RT "Arabidopsis thaliana chromosome 1 BAC T518 sequence."
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC007060; AAD25743.1; -;
 DR InterPro; IPR007114; MFS.
 DR PROSITE; PS50850; MFS; 1.
 SQ SEQUENCE 510 AA; 55089 MW; 0DC63CC41F4712F8 CRC64;

Query Match 62.5%; Score 35; DB 10; Length 510;
 Best Local Similarity 60.0%; Pred. No. 1.2e+02;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 BEVVPXGMHY 10
 |||||
 Db 12 BEVKKPGIHF 21

Search completed: December 22, 2003, 17:51:29

Job time : 25.2667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:37:03 ; Search time 10.1333 Seconds
(without alignments)
45.930 Million cell updates/sec

Title: US-09-909-164-12

Perfect score: 54

Sequence: 1 EVVPGMDYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgm2_6/ptodata/1/iaa/5A COMB.pcp.*
2: /cgm2_6/ptodata/1/iaa/5B COMB.pcp.*
3: /cgm2_6/ptodata/1/iaa/6A COMB.pcp.*
4: /cgm2_6/ptodata/1/iaa/6B COMB.pcp.*
5: /cgm2_6/ptodata/1/iaa/PTCUS COMB.pcp.*
6: /cgm2_6/ptodata/1/iaa/backfile1.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|----------------------|
| 1 | 36 | 66.7 | 341 | 3 | US-08-853-948B-4 |
| 2 | 36 | 66.7 | 348 | 3 | US-08-853-948B-5 |
| 3 | 36 | 66.7 | 368 | 4 | US-09-697-367-24 |
| 4 | 36 | 66.7 | 1049 | 4 | US-09-394-272-10 |
| 5 | 36 | 66.7 | 1068 | 2 | US-08-429-054A-11 |
| 6 | 36 | 66.7 | 1068 | 2 | US-08-718-777-7 |
| 7 | 36 | 66.7 | 1068 | 3 | US-09-051-341-7 |
| 8 | 36 | 66.7 | 1068 | 4 | US-09-394-272-8 |
| 9 | 36 | 66.7 | 1081 | 4 | US-09-394-272-4 |
| 10 | 36 | 66.7 | 1083 | 4 | US-09-394-272-11 |
| 11 | 36 | 66.7 | 1084 | 4 | US-09-394-272-9 |
| 12 | 34 | 63.0 | 140 | 3 | US-08-569-147-76 |
| 13 | 34 | 63.0 | 140 | 3 | US-08-569-147-82 |
| 14 | 34 | 63.0 | 1065 | 4 | US-09-252-991A-31637 |
| 15 | 33 | 61.1 | 59 | 4 | US-08-963-851-14 |
| 16 | 33 | 61.1 | 378 | 1 | US-08-070-165P-8 |
| 17 | 33 | 61.1 | 378 | 2 | US-08-885-418-8 |
| 18 | 33 | 61.1 | 473 | 4 | US-09-252-991A-26805 |
| 19 | 33 | 61.1 | 765 | 4 | US-09-819-989-4 |
| 20 | 33 | 61.1 | 801 | 3 | US-09-383-630-6 |
| 21 | 33 | 61.1 | 811 | 4 | US-09-819-989-2 |
| 22 | 33 | 61.1 | 883 | 4 | US-09-667-373-4 |
| 23 | 32 | 59.3 | 65 | 6 | 5177197-51 |
| 24 | 32 | 59.3 | 102 | 2 | US-08-580-988A-23 |
| 25 | 32 | 59.3 | 152 | 2 | US-08-460-694-4 |
| 26 | 32 | 59.3 | 152 | 3 | US-08-460-744-4 |
| 27 | 32 | 59.3 | 152 | 3 | US-07-667-711B-4 |

Sequence 7, Appli
Sequence 21, Appli
Sequence 21, Appli
Sequence 21, Appli
Sequence 20, Appli
Sequence 22, Appli
Sequence 22, Appli
Sequence 22, Appli
Sequence 21, Appli
Sequence 4172, Ap
Sequence 6, Appli
Sequence 6, Appli
Sequence 4, Appli
Sequence 6, Appli
Sequence 23, Appli

1 US-08-193-977-7
2 US-08-464-517-21
3 US-08-246-361A-21
4 US-08-463-772-21
5 PCT-US93-05000-21
3 US-08-926-842B-20
2 US-08-464-517-22
3 US-08-246-361A-22
3 US-08-463-772-22
5 PCT-US93-05000-22
4 US-08-926-842B-21
4 US-09-107-532A-4172
3 US-08-464-517-6
3 US-08-463-772-6
2 US-08-246-361A-4
5 PCT-US93-05000-4
5 PCT-US93-05000-6
2 US-08-464-517-23

ALIGNMENTS

RESULT 1:
US-08-853-948B-4
; Sequence 4, Application US/08853948B
; Patent No. 6210943
; GENERAL INFORMATION:
; APPLICANT: AKIHAMA, Toyota
; TITLE OF INVENTION: SUCROSE PHOSPHATE SYNTHASE FROM CITRUS AND DNA ENCODING
; FILE REFERENCE: 0049-0235-0
; CURRENT APPLICATION NUMBER: US/08/853,948B
; CURRENT FILING DATE: 1997-05-09
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Citrus unshiu
; FEATURE:
; OTHER INFORMATION: Xaa at position 109 is one of Ala, Arg, Asn, Asp,
; OTHER INFORMATION: Cys, Gln, Glu, Gly, His, Ile, Leu, Lys, Met, Phe,
; OTHER INFORMATION: Pro, Ser, Thr, Trp, Tyr, or Val
US-08-853-948B-4

Query Match 66.7%; Score 36; DB 3; Length 341;
Best Local Similarity 66.7%; Pred. No. 18;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 3 VVPXGMDYS 11
DB 228 VPPGMDPS 236

RESULT 2
US-08-853-948B-5
; Sequence 5, Application US/08853948B
; Patent No. 6210943
; GENERAL INFORMATION:
; APPLICANT: AKIHAMA, Toyota
; TITLE OF INVENTION: SUCROSE PHOSPHATE SYNTHASE FROM CITRUS AND DNA ENCODING
; FILE REFERENCE: 0049-0235-0
; CURRENT APPLICATION NUMBER: US/08/853,948B
; CURRENT FILING DATE: 1997-05-09
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 348
; TYPE: PRT
; ORGANISM: Citrus unshiu
US-08-853-948B-5

Query Match 66.7%; Score 36; DB 3; Length 348;
Best Local Similarity 66.7%; Pred. No. 19;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
|:| |:| |:|
Db 234 VIPPGMDFS 242

RESULT 3

US-09-697-367-24
; Sequence 24, Application US/09697367
; Patent No. 6323015
; GENERAL INFORMATION:
; APPLICANT: Orozco Jr., Emil M.
; APPLICANT: Calimi, Perry G.
; APPLICANT: Weng, Zude
; APPLICANT: Tarczynski, Mitchell
; TITLE OF INVENTION: SUCROSE PHOSPHATE SYNTHASE
; FILE REFERENCE: BB1166 US NA
; CURRENT APPLICATION NUMBER: US/09/697,367
; CURRENT FILING DATE: 2000-10-26
; PRIOR APPLICATION NUMBER: 60/084,529
; PRIOR FILING DATE: 1998-MAY-07
; PRIOR APPLICATION NUMBER: PCT/US99/09865
; PRIOR FILING DATE: 1999-MAY-06
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 24
; LENGTH: 368
; TYPE: PRT
; ORGANISM: Zea mays
US-09-697-367-24

Query Match 66.7%; Score 36; DB 4; Length 368;
Best Local Similarity 66.7%; Pred. No. 20;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
|:| |:| |:|
Db 217 VIPPGMDFS 225

RESULT 4

US-09-394-272-10
; Sequence 10, Application US/09394272
; Patent No. 6472588
; GENERAL INFORMATION:
; APPLICANT: Haigler, Candace H.
; APPLICANT: Holaday, A. Scott
; TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
; FILE REFERENCE: 201304/1000
; CURRENT APPLICATION NUMBER: US/09/394,272
; CURRENT FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1049
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-394-272-10

Query Match 66.7%; Score 36; DB 4; Length 1049;
Best Local Similarity 66.7%; Pred. No. 67;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
|:| |:| |:|
Db 436 VIPPGMDFS 444

RESULT 5

US-08-429-054A-11
; Sequence 11, Application US/08429054A
; Patent No. 5917126
; GENERAL INFORMATION:
; APPLICANT: VAN ASSCHE, CHARLES; LANDO, DANIELLE; BRUNEAU,
; APPLICANT: JEAN; VOELKER, TONI; GERVAIS, MONICA
; TITLE OF INVENTION: SUCROSE PHOSPHATE SYNTHETASE (SPS),
; TITLE OF INVENTION: A PREPARATION METHOD AND CDNA THEREFOR, AND USE OF THE
; TITLE OF INVENTION: CDNA FOR MODIFYING SPS EXPRESSION IN PLANT CELLS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIERMAN AND MUSERLIAN
; STREET: 600 THIRD AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10016

COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/429,054A
; FILING DATE: 26-APR-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 842,337
; FILING DATE: 20-March-1992
; APPLICATION NUMBER: PCT/FR 91/00593
; FILING DATE: 18-July-1991
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: French 90402094.9
; FILING DATE: 20-July-1990
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Charles A. Muserlian
; REGISTRATION NUMBER: 19,683
; REFERENCE/DOCKET NUMBER: 146.1137
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 661-8000
; TELEFAX: (212) 661-8002
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1068
; TYPE: Amino acid
; STRANDEDNESS: Single
; TOPOLOGY: Unknown
; MOLECULE TYPE: Peptide
US-08-429-054A-11

Query Match 66.7%; Score 36; DB 2; Length 1068;
Best Local Similarity 66.7%; Pred. No. 68;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
|:| |:| |:|
Db 435 VIPPGMDFS 443

RESULT 6

US-08-718-777-7
; Sequence 7, Application US/08718777
; Patent No. 5981852
; GENERAL INFORMATION:
; APPLICANT: Van Assche, C.
; APPLICANT: Lando, D.
; APPLICANT: Bruneau, J. M.
; APPLICANT: Voelker, T.
; APPLICANT: Gervais, M.
; TITLE OF INVENTION: MODIFICATION OF SUCROSE

;/ TITLE OF INVENTION: PHOSPHATE
;/ TITLE OF INVENTION: SYNTHASE IN PLANTS
;/ NUMBER OF SEQUENCES: 14
;/ CORRESPONDENCE ADDRESS:
;/ ADDRESSEE: Law Offices of Barbara Rae-Venter
;/ STREET: 260 Sheridan Avenue, Suite 440
;/ CITY: Palo Alto
;/ STATE: California
;/ COUNTRY: USA
;/ ZIP: 94306
;/ COMPUTER READABLE FORM:
;/ MEDIUM TYPE: Floppy disk
;/ COMPUTER: IBM PC compatible
;/ OPERATING SYSTEM: PC-DOS/MS-DOS
;/ SOFTWARE: PatentIn Release #1.0, Version #1.30
;/ CURRENT APPLICATION DATA:
;/ APPLICATION NUMBER: US/08/718,777
;/ FILING DATE: NOT YET ASSIGNED
;/ CLASSIFICATION: 800
;/ PRIOR APPLICATION DATA:
;/ APPLICATION NUMBER: US 08/175,471
;/ FILING DATE: 27-DEC-1993
;/ ATTORNEY/AGENT INFORMATION:
;/ NAME: Barbara Rae-Venter
;/ REGISTRATION NUMBER: 32,750
;/ REFERENCE/DOCKET NUMBER: CGNE.072.02US
;/ TELEPHONE: (415)328-4400
;/ TELEFAX: (415)328-4477
;/ INFORMATION FOR SEQ ID NO: 7:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 1068 amino acids
;/ TYPE: amino acid
;/ TOPOLOGY: linear
;/ MOLECULE TYPE: protein
;/ US-08-718-777-7

Query Match 66.7%; Score 36; DB 2; Length 1068;
Best Local Similarity 66.7%; Pred. No. 68;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
|:|:|:|:
Db 435 VIPPGMDFS 443

RESULT 7
US-09-051-341-7
;/ Sequence 7, Application US/09051341
;/ Patent No. 6124528
;/ GENERAL INFORMATION:
;/ APPLICANT: Shewmaker, C. K.
;/ TITLE OF INVENTION: MODIFICATION OF SOLUBLE SOLIDS USING
;/ TITLE OF INVENTION: SUCROSE PHOSPHATE SYNTHASE ENCODING SEQUENCE
;/ NUMBER OF SEQUENCES: 14
;/ CORRESPONDENCE ADDRESS:
;/ ADDRESSEE: Rae-Venter Law Group, P.C.
;/ STREET: 260 Sheridan Avenue, Suite 440
;/ CITY: Palo Alto
;/ STATE: California
;/ COUNTRY: USA
;/ ZIP: 94306
;/ COMPUTER READABLE FORM:
;/ MEDIUM TYPE: Floppy disk
;/ COMPUTER: IBM PC compatible
;/ OPERATING SYSTEM: PC-DOS/MS-DOS
;/ SOFTWARE: PatentIn Release #1.0, Version #1.30
;/ CURRENT APPLICATION DATA:
;/ APPLICATION NUMBER: US/09/051,341
;/ FILING DATE:
;/ CLASSIFICATION: 800
;/ PRIOR APPLICATION DATA:
;/ APPLICATION NUMBER: PCT/US96/17351

;/ FILING DATE: 25-OCT-1996
;/ APPLICATION NUMBER: US 08/549,016
;/ FILING DATE: 27-OCT-1995
;/ PRIOR APPLICATION DATA:
;/ APPLICATION NUMBER: US 08/372,200
;/ FILING DATE: 12-JAN-1995
;/ ATTORNEY/AGENT INFORMATION:
;/ NAME: Barbara Rae-Venter, Ph.D.,
;/ REGISTRATION NUMBER: 32,750
;/ REFERENCE/DOCKET NUMBER: CGNE.110.02US
;/ TELECOMMUNICATION INFORMATION:
;/ TELEPHONE: (415)328-4400
;/ TELEFAX: (415)328-4477
;/ INFORMATION FOR SEQ ID NO: 7:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 1068 amino acids
;/ TYPE: amino acid
;/ TOPOLOGY: linear
;/ MOLECULE TYPE: protein
;/ US-09-051-341-7

Query Match 66.7%; Score 36; DB 3; Length 1068;
Best Local Similarity 66.7%; Pred. No. 68;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
|:|:|:|:
Db 435 VIPPGMDFS 443

RESULT 8
US-09-394-272-8
;/ Sequence 8, Application US/09394272
;/ Patent No. 6472588
;/ GENERAL INFORMATION:
;/ APPLICANT: Haigler, Candace H.
;/ APPLICANT: Holaday, A. Scott
;/ TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
;/ TITLE OF INVENTION: EXPRESSION OF SUCROSE PHOSPHATE SYNTHASE
;/ FILE REFERENCE: 201304/1000
;/ CURRENT APPLICATION NUMBER: US/09/394,272
;/ CURRENT FILING DATE: 1999-09-10
;/ NUMBER OF SEQ ID NOS: 14
;/ SOFTWARE: PatentIn Ver. 2.0
;/ SEQ ID NO 8
;/ LENGTH: 1068
;/ TYPE: PRT
;/ ORGANISM: Zea mays
;/ US-09-394-272-8

Query Match 66.7%; Score 36; DB 4; Length 1068;
Best Local Similarity 66.7%; Pred. No. 68;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
|:|:|:|:
Db 435 VIPPGMDFS 443

RESULT 9
US-09-394-272-4
;/ Sequence 4, Application US/09394272
;/ Patent No. 6472588
;/ GENERAL INFORMATION:
;/ APPLICANT: Haigler, Candace H.
;/ APPLICANT: Holaday, A. Scott
;/ TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
;/ TITLE OF INVENTION: EXPRESSION OF SUCROSE PHOSPHATE SYNTHASE
;/ FILE REFERENCE: 201304/1000
;/ CURRENT APPLICATION NUMBER: US/09/394,272
;/ CURRENT FILING DATE: 1999-09-10
;/ NUMBER OF SEQ ID NOS: 14
;/ SOFTWARE: PatentIn Ver. 2.0

```
; SEQ ID NO 4
; LENGTH: 1081
; TYPE: PRT
; ORGANISM: Craterostigma plantagineum
US-09-394-272-4

Query Match          66.7%; Score 36; DB 4; Length 1081;
Best Local Similarity 66.7%; Pred. No. 69;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      3 VVPXGMDYS 11
Db      445 VIPPGMDFS 453

RESULT 10
US-09-394-272-11
; Sequence 11, Application US/09394272
; Patent No. 6472588
; GENERAL INFORMATION:
; APPLICANT: Haigler, Candace H.
; APPLICANT: Holaday, A. Scott
; TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
; TITLE OF INVENTION: EXPRESSION OF SUCROSE PHOSPHATE SYNTHASE
; FILE REFERENCE: 201304/1000
; CURRENT APPLICATION NUMBER: US/09/394,272
; CURRENT FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 1083
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-394-272-11

Query Match          66.7%; Score 36; DB 4; Length 1083;
Best Local Similarity 66.7%; Pred. No. 69;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      3 VVPXGMDYS 11
Db      483 VIPPGMDFS 491

RESULT 11
US-09-394-272-9
; Sequence 9, Application US/09394272
; Patent No. 6472588
; GENERAL INFORMATION:
; APPLICANT: Haigler, Candace H.
; APPLICANT: Holaday, A. Scott
; TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
; TITLE OF INVENTION: EXPRESSION OF SUCROSE PHOSPHATE SYNTHASE
; FILE REFERENCE: 201304/1000
; CURRENT APPLICATION NUMBER: US/09/394,272
; CURRENT FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1084
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-394-272-9

Query Match          66.7%; Score 36; DB 4; Length 1084;
Best Local Similarity 66.7%; Pred. No. 69;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      3 VVPXGMDYS 11
Db      453 VIPPGMDFS 461

; SEQ ID NO 4
; LENGTH: 1081
; TYPE: PRT
; ORGANISM: Craterostigma plantagineum
US-09-394-272-4

Query Match          66.7%; Score 36; DB 4; Length 1081;
Best Local Similarity 66.7%; Pred. No. 69;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      3 VVPXGMDYS 11
Db      445 VIPPGMDFS 453

RESULT 12
US-08-569-147-76
; Sequence 76, Application US/08569147
; Patent No. 6180377
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: HUMANISED ANTIBODIES
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
; ADDRESSEE: No. 6180377ris, LLP
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/569,147
; FILING DATE: 25-March-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Tutjillo, Doreen Yanko
; REGISTRATION NUMBER: 35,719
; REFERENCE/DOCKET NUMBER: CARP-0047
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 76:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 140 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-569-147-76

Query Match          63.0%; Score 34; DB 3; Length 140;
Best Local Similarity 75.0%; Pred. No. 16;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      3 VVPXGMDY 10
Db      122 VVPTGFDY 129

RESULT 13
US-08-569-147-82
; Sequence 82, Application US/08569147
; Patent No. 6180377
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: HUMANISED ANTIBODIES
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
; ADDRESSEE: No. 6180377ris, LLP
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/569,147
; FILING DATE: 25-March-1996
```

```
/ CLASSIFICATION: 536
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Trujillo, Doreen Yanko
/ REGISTRATION NUMBER: 35,719
/ REFERENCE/DOCKET NUMBER: CARP-0047
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (215) 568-3100
/ TELEFAX: (215) 568-3439
/ INFORMATION FOR SEQ ID NO: 82:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 140 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-569-147-82

Query Match      63.0%; Score 34; DB 3; Length 140;
Best Local Similarity 75.0%; Pred. No. 16;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      3 VVPXGMDY 10
      |||||
Db      122 VVPTGFDY 129

RESULT 14
US-09-252-991A-31637
/ Sequence 31637, Application US/09252991A
/ Patent No. 6551795
/ GENERAL INFORMATION:
/ APPLICANT: Marc J. Rubenfield et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
/ FILE REFERENCE: 107196.136
/ CURRENT APPLICATION NUMBER: US/09/252,991A
/ CURRENT FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074,788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094,190
/ PRIOR FILING DATE: 1998-07-27
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 31637
/ LENGTH: 1065
/ TYPE: PRT
/ ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31637
```

```
Query Match      63.0%; Score 34; DB 4; Length 1065;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      5 PXGMDYS 11
      |||||
Db      324 POGMDYS 330
```

```
RESULT 15
US-08-963-851-14
/ Sequence 14, Application US/08963851
/ Patent No. 6300116
/ GENERAL INFORMATION:
/ APPLICANT: VAN DER OSTEN, CLAUS
/ APPLICANT: HALKIER, TORDEN
/ APPLICANT: ANDERSEN, CARSTEN
/ APPLICANT: BAUDITZ, PETER
/ APPLICANT: HANSEN, PETER KAMP
/ TITLE OF INVENTION: PROTEASE VARIANTS AND COMPOSITIONS
/ FILE REFERENCE: 4946,200-US
/ CURRENT APPLICATION NUMBER: US/08/963,851
/ CURRENT FILING DATE: 1997-11-04
/ NUMBER OF SEQ ID NOS: 35
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 14
```

```
/ LENGTH: 59
/ TYPE: PRT
/ ORGANISM: Enterococcus faecalis
/ US-08-963-851-14

Query Match      61.1%; Score 33; DB 4; Length 59;
Best Local Similarity 45.5%; Pred. No. 9.6;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY      1 BEVVPXGMDYS 11
      |.:|.:|
Db      38 EXHIFGLEYYS 48

Search completed: December 22, 2003, 16:43:43
Job time : 10.1333 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:40:14 ; Search time 20.8667 Seconds
(without alignments)
98.451 Million cell updates/sec

Title: US-09-909-164-12
Perfect score: 54
Sequence: 1 EEVFXGMDYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 696363 seqs, 186758610 residues

Total number of hits satisfying chosen parameters: 696363

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------------|
| 1 | 36 | 66.7 | 126 | 12 | US-10-393-840-946 |
| 2 | 36 | 66.7 | 937 | 12 | Sequence 946, App |
| 3 | 36 | 66.7 | 938 | 12 | Sequence 73, Appl |
| 4 | 36 | 66.7 | 1049 | 15 | Sequence 168, App |
| 5 | 36 | 66.7 | 1062 | 12 | Sequence 10, Appl |
| 6 | 36 | 66.7 | 1068 | 15 | Sequence 71, Appl |
| 7 | 36 | 66.7 | 1081 | 15 | Sequence 8, Appl |
| 8 | 36 | 66.7 | 1083 | 15 | Sequence 4, Appl |
| 9 | 36 | 66.7 | 1084 | 15 | Sequence 11, Appl |
| 10 | 35 | 64.8 | 440 | 11 | Sequence 27, Appl |
| 11 | 34.5 | 63.9 | 588 | 12 | Sequence 1660, Ap |
| 12 | 34 | 63.0 | 222 | 12 | Sequence 58, Appl |
| 13 | 34 | 63.0 | 1062 | 9 | Sequence 511, Ap |
| 14 | 34 | 63.0 | 3472 | 12 | Sequence 4, Appl |
| 15 | 34 | 63.0 | 3472 | 14 | Sequence 4, Appl |

| | | | | | | |
|----|----|------|------|----|---------------------|-------------------|
| 16 | 34 | 63.0 | 3472 | 14 | US-10-034-623-4 | Sequence 4, Appli |
| 17 | 34 | 63.0 | 3472 | 15 | US-10-027-801-4 | Sequence 4, Appli |
| 18 | 33 | 61.1 | 59 | 10 | US-09-948-080-14 | Sequence 14, Appl |
| 19 | 33 | 61.1 | 236 | 12 | US-10-029-386-32076 | Sequence 32076, A |
| 20 | 33 | 61.1 | 283 | 10 | US-09-738-626-4881 | Sequence 4881, Ap |
| 21 | 33 | 61.1 | 299 | 9 | US-09-815-242-10697 | Sequence 10697, A |
| 22 | 33 | 61.1 | 322 | 15 | US-10-156-761-12418 | Sequence 12418, A |
| 23 | 33 | 61.1 | 736 | 10 | US-09-978-295A-526 | Sequence 526, App |
| 24 | 33 | 61.1 | 736 | 10 | US-09-978-697-526 | Sequence 526, App |
| 25 | 33 | 61.1 | 736 | 10 | US-09-978-192A-526 | Sequence 526, App |
| 26 | 33 | 61.1 | 736 | 10 | US-09-999-832A-526 | Sequence 526, App |
| 27 | 33 | 61.1 | 736 | 11 | US-09-978-189-526 | Sequence 526, App |
| 28 | 33 | 61.1 | 736 | 11 | US-09-978-608A-526 | Sequence 526, App |
| 29 | 33 | 61.1 | 736 | 11 | US-09-978-585A-526 | Sequence 526, App |
| 30 | 33 | 61.1 | 736 | 11 | US-09-978-191A-526 | Sequence 526, App |
| 31 | 33 | 61.1 | 736 | 11 | US-09-978-403A-526 | Sequence 526, App |
| 32 | 33 | 61.1 | 736 | 11 | US-09-978-564A-526 | Sequence 526, App |
| 33 | 33 | 61.1 | 736 | 11 | US-09-999-833A-526 | Sequence 526, App |
| 34 | 33 | 61.1 | 736 | 11 | US-09-981-915A-526 | Sequence 526, App |
| 35 | 33 | 61.1 | 736 | 11 | US-09-978-824-526 | Sequence 526, App |
| 36 | 33 | 61.1 | 736 | 11 | US-09-918-585A-526 | Sequence 526, App |
| 37 | 33 | 61.1 | 736 | 11 | US-09-978-423A-526 | Sequence 526, App |
| 38 | 33 | 61.1 | 736 | 11 | US-09-978-193A-526 | Sequence 526, App |
| 39 | 33 | 61.1 | 736 | 11 | US-09-999-830A-526 | Sequence 526, App |
| 40 | 33 | 61.1 | 736 | 11 | US-09-978-757A-526 | Sequence 526, App |
| 41 | 33 | 61.1 | 736 | 11 | US-09-978-187B-526 | Sequence 526, App |
| 42 | 33 | 61.1 | 736 | 11 | US-09-978-643A-526 | Sequence 526, App |
| 43 | 33 | 61.1 | 736 | 12 | US-09-978-375A-526 | Sequence 526, App |
| 44 | 33 | 61.1 | 736 | 12 | US-09-978-188A-526 | Sequence 526, App |
| 45 | 33 | 61.1 | 736 | 12 | US-09-978-298A-526 | Sequence 526, App |

ALIGNMENTS

RESULT 1
US-10-393-840-946
; Sequence 946, Application US/10393840
; Publication No. US2003022922A1
; GENERAL INFORMATION:
; APPLICANT: Bloksberg, Leonard N.
; TITLE OF INVENTION: Materials and Methods for the
; FILE OF INVENTION: Modification of Plant Cell Wall Polysaccharides
; FILE REFERENCE: 11000.1012c3
; CURRENT APPLICATION NUMBER: US/10/393,840
; CURRENT FILING DATE: 2003-03-20
; PRIOR APPLICATION NUMBER: US 09/636,800
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: US 09/170,862
; PRIOR FILING DATE: 1998-10-13
; PRIOR APPLICATION NUMBER: US 60/148,426
; PRIOR FILING DATE: 1999-08-11
; PRIOR APPLICATION NUMBER: PCT NZ/99/00169
; NUMBER OF SEQ ID NOS: 956
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 946
; LENGTH: 126
; TYPE: PRT
; ORGANISM: Pinus radiata
US-10-393-840-946

Query Match 66.7%; Score 36; DB 12; Length 126;
Best Local Similarity 66.7%; Pred. No. 14;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVFXGMDYS 11
|:|:|:|:
Db 23 VVFXGMDYS 31

RESULT 2
US-10-289-757-73

```

; Sequence 73, Application US/10289757
; Publication No. US20030180751A1
; GENERAL INFORMATION:
; APPLICANT: Demmer, Jeroen
; APPLICANT: Forster, Richard L
; APPLICANT: Gibson, John Bryan
; APPLICANT: Shenk, Michael Andrew
; APPLICANT: No. US20030180751A1riss, Geoffrey
; APPLICANT: Glenn, Matthew
; APPLICANT: Saulsbury, Keith Martin
; APPLICANT: Hall, Claire
; TITLE OF INVENTION: Compositions isolated from forage
; TITLE OF INVENTION: grasses and methods for their use
; FILE REFERENCE: 11000.1061U
; CURRENT APPLICATION NUMBER: US/10/289,757
; PRIOR FILING DATE: 2002-11-07
; PRIOR FILING DATE: 2001-11-07
; NUMBER OF SEQ ID NOS: 218
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 73
; LENGTH: 937
; TYPE: PRT
; ORGANISM: Lolium perenne
US-10-289-757-73

```

Query Match 66.7%; Score 36; DB 12; Length 937;

Best Local Similarity 66.7%; Pred. No. 1.3e+02; Indels 0; Gaps 0;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
|:|:|:|:|

Db 462 VIPPGMDFS 470

RESULT 3

```

US-10-289-757-168
; Sequence 168, Application US/10289757
; Publication No. US20030180751A1
; GENERAL INFORMATION:
; APPLICANT: Demmer, Jeroen
; APPLICANT: Forster, Richard L
; APPLICANT: Gibson, John Bryan
; APPLICANT: Shenk, Michael Andrew
; APPLICANT: No. US20030180751A1riss, Geoffrey
; APPLICANT: Glenn, Matthew
; APPLICANT: Saulsbury, Keith Martin
; APPLICANT: Hall, Claire
; TITLE OF INVENTION: Compositions isolated from forage
; TITLE OF INVENTION: grasses and methods for their use
; FILE REFERENCE: 11000.1061U
; CURRENT APPLICATION NUMBER: US/10/289,757
; PRIOR FILING DATE: 2002-11-07
; PRIOR FILING DATE: 2001-11-07
; NUMBER OF SEQ ID NOS: 218
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 168
; LENGTH: 938
; TYPE: PRT
; ORGANISM: Lolium perenne
US-10-289-757-168

```

Query Match 66.7%; Score 36; DB 12; Length 938;

Best Local Similarity 66.7%; Pred. No. 1.3e+02; Indels 0; Gaps 0;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
|:|:|:|:|

Db 462 VIPPGMDFS 470

RESULT 4

```

US-10-289-757-168
; Sequence 168, Application US/10289757
; Publication No. US20030180751A1
; GENERAL INFORMATION:
; APPLICANT: Demmer, Jeroen
; APPLICANT: Forster, Richard L
; APPLICANT: Gibson, John Bryan
; APPLICANT: Shenk, Michael Andrew
; APPLICANT: No. US20030180751A1riss, Geoffrey
; APPLICANT: Glenn, Matthew
; APPLICANT: Saulsbury, Keith Martin
; APPLICANT: Hall, Claire
; TITLE OF INVENTION: Compositions isolated from forage
; TITLE OF INVENTION: grasses and methods for their use
; FILE REFERENCE: 11000.1061U
; CURRENT APPLICATION NUMBER: US/10/289,757
; PRIOR FILING DATE: 2002-11-07
; PRIOR FILING DATE: 2001-11-07
; NUMBER OF SEQ ID NOS: 218
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 168
; LENGTH: 938
; TYPE: PRT
; ORGANISM: Lolium perenne
US-10-289-757-168

```

US-10-217-700-10

```

; Sequence 10, Application US/10217700
; Publication No. US20030070191A1
; GENERAL INFORMATION:
; APPLICANT: Haigler, Candace H.
; APPLICANT: Holaday, A. Scott
; TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
; TITLE OF INVENTION: EXPRESSION OF SUCROSE PHOSPHATE SYNTHASE
; FILE REFERENCE: 201304/1000
; CURRENT APPLICATION NUMBER: US/10/217,700
; CURRENT FILING DATE: 2002-08-12
; EARLIER APPLICATION NUMBER: 09/394,272
; PRIOR FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1049
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-217-700-10

```

Query Match 66.7%; Score 36; DB 15; Length 1049;

Best Local Similarity 66.7%; Pred. No. 1.5e+02; Indels 0; Gaps 0;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
|:|:|:|:|

Db 436 VIPPGMDFS 444

RESULT 5

```

US-10-289-757-71
; Sequence 71, Application US/10289757
; Publication No. US20030180751A1
; GENERAL INFORMATION:
; APPLICANT: Demmer, Jeroen
; APPLICANT: Forster, Richard L
; APPLICANT: Gibson, John Bryan
; APPLICANT: Shenk, Michael Andrew
; APPLICANT: No. US20030180751A1riss, Geoffrey
; APPLICANT: Glenn, Matthew
; APPLICANT: Saulsbury, Keith Martin
; APPLICANT: Hall, Claire
; TITLE OF INVENTION: Compositions isolated from forage
; TITLE OF INVENTION: grasses and methods for their use
; FILE REFERENCE: 11000.1061U
; CURRENT APPLICATION NUMBER: US/10/289,757
; PRIOR FILING DATE: 2002-11-07
; PRIOR FILING DATE: 2001-11-07
; NUMBER OF SEQ ID NOS: 218
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 71
; LENGTH: 1062
; TYPE: PRT
; ORGANISM: Festuca arundinacea
US-10-289-757-71

```

Query Match 66.7%; Score 36; DB 12; Length 1062;

Best Local Similarity 66.7%; Pred. No. 1.5e+02; Indels 0; Gaps 0;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
|:|:|:|:|

Db 466 VIPPGMDFS 474

RESULT 6

```

US-10-217-700-8
; Sequence 8, Application US/10217700
; Publication No. US20030070191A1
; GENERAL INFORMATION:
; APPLICANT: Haigler, Candace H.

```

```

; APPLICANT: Holaday, A. Scott
; TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
; TITLE OF INVENTION: EXPRESSION OF SUCROSE PHOSPHATE SYNTHASE
; FILE REFERENCE: 201304/1000
; CURRENT APPLICATION NUMBER: US/10/217,700
; CURRENT FILING DATE: 2002-08-12
; EARLIER APPLICATION NUMBER: 09/394,272
; EARLIER FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1068
; TYPE: PRT
; ORGANISM: Zea mays
US-10-217-700-8

Query Match      66.7%; Score 36; DB 15; Length 1068;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      3 VVPXGMDYS 11
Db      435 VIPPGMDFS 443

RESULT 7
US-10-217-700-4
; Sequence 4, Application US/10217700
; Publication No. US20030070191A1
; GENERAL INFORMATION:
; APPLICANT: Haigler, Candace H.
; APPLICANT: Holaday, A. Scott
; TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
; TITLE OF INVENTION: EXPRESSION OF SUCROSE PHOSPHATE SYNTHASE
; FILE REFERENCE: 201304/1000
; CURRENT APPLICATION NUMBER: US/10/217,700
; CURRENT FILING DATE: 2002-08-12
; EARLIER APPLICATION NUMBER: 09/394,272
; EARLIER FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1081
; TYPE: PRT
; ORGANISM: Craterostigma plantagineum
US-10-217-700-4

Query Match      66.7%; Score 36; DB 15; Length 1081;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      3 VVPXGMDYS 11
Db      445 VIPPGMDFS 453

RESULT 8
US-10-217-700-11
; Sequence 11, Application US/10217700
; Publication No. US20030070191A1
; GENERAL INFORMATION:
; APPLICANT: Haigler, Candace H.
; APPLICANT: Holaday, A. Scott
; TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
; TITLE OF INVENTION: EXPRESSION OF SUCROSE PHOSPHATE SYNTHASE
; FILE REFERENCE: 201304/1000
; CURRENT APPLICATION NUMBER: US/10/217,700
; CURRENT FILING DATE: 2002-08-12
; EARLIER APPLICATION NUMBER: 09/394,272
; EARLIER FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11

```

```

; LENGTH: 1083
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-217-700-11

Query Match      66.7%; Score 36; DB 15; Length 1083;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      3 VVPXGMDYS 11
Db      483 VIPPGMDFS 491

RESULT 9
US-10-217-700-9
; Sequence 9, Application US/10217700
; Publication No. US20030070191A1
; GENERAL INFORMATION:
; APPLICANT: Haigler, Candace H.
; APPLICANT: Holaday, A. Scott
; TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
; TITLE OF INVENTION: EXPRESSION OF SUCROSE PHOSPHATE SYNTHASE
; FILE REFERENCE: 201304/1000
; CURRENT APPLICATION NUMBER: US/10/217,700
; CURRENT FILING DATE: 2002-08-12
; EARLIER APPLICATION NUMBER: 09/394,272
; EARLIER FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1084
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-217-700-9

Query Match      66.7%; Score 36; DB 15; Length 1084;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      3 VVPXGMDYS 11
Db      453 VIPPGMDFS 461

RESULT 10
US-09-813-408-27
; Sequence 27, Application US/09813408
; Publication No. US20030049619A1
; GENERAL INFORMATION:
; APPLICANT: Delagrave, Simon
; APPLICANT: Marrs, Barry
; TITLE OF INVENTION: Methods For The Synthesis Of Polynucleotides And Combinatorial
; TITLE OF INVENTION: Of Polynucleotides
; FILE REFERENCE: HER0041
; CURRENT APPLICATION NUMBER: US/09/813,408
; CURRENT FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 27
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Aeropyrum pernix
US-09-813-408-27

Query Match      64.8%; Score 35; DB 11; Length 440;
Best Local Similarity 66.7%; Pred. No. 92;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      2 EVVFXGMDY 10
Db      120 EVLPWGVY 128

```

```
RESULT 11
US-10-094-749-1660
; Sequence 1660, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOTIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOKYU
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1660
; LENGTH: 588
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-1660

Query Match      63.9%; Score 34.5; DB 12; Length 588;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 8; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

QY      1 EEVVPXGMDY 10
DB      127 EEVVP-GMDF 135

RESULT 12
US-10-091-007-58
; Sequence 58, Application US/10091007
; Publication No. US20030170782A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics limited
; APPLICANT: Le Page, Richard W F
; APPLICANT: Hanniffy, Sean B
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21978WO
; CURRENT APPLICATION NUMBER: US/10/091,007
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: GB 9921125.2
; PRIOR FILING DATE: 1999-09-07
; NUMBER OF SEQ ID NOS: 276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 58
; LENGTH: 222
; TYPE: PRT
; ORGANISM: Streptococcus agalactiae
US-10-091-007-58

Query Match      63.0%; Score 34; DB 12; Length 222;
Best Local Similarity 50.0%; Pred. No. 68;

QY      1 EEVVPXGMDY 10
DB      127 EEVVP-GMDF 135

RESULT 13
US-09-815-242-5111
; Sequence 5111, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zvakind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: EUTRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5111
; LENGTH: 1062
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-5111

Query Match      63.0%; Score 34; DB 9; Length 1062;
Best Local Similarity 85.7%; Pred. No. 3.9e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      5 PXGMDYS 11
DB      321 PQGMDYS 327

RESULT 14
US-10-029-120-4
; Sequence 4, Application US/10029120
; Publication No. US20030175708A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCRP.002A
; CURRENT APPLICATION NUMBER: US/10/029,120
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
```



```
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-029-120-4

Query Match      63.0%; Score 34; DB 12; Length 3472;
Best Local Similarity 45.5%; Pred. No. 1.5e+03;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY      1 EEVVPXGMDYS 11
Db      2294 EDVIPRGISFS 2304

RESULT 15
US-10-027-806-4
; Sequence 4, Application US/10027806
; Publication No. US20020160476A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCCRP.002A
; CURRENT APPLICATION NUMBER: US/10/027,806
; CURRENT FILING DATE: 2001-12-21
; PRIOR FILING DATE: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-027-806-4

Query Match      63.0%; Score 34; DB 14; Length 3472;
Best Local Similarity 45.5%; Pred. No. 1.5e+03;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY      1 EEVVPXGMDYS 11
Db      2294 EDVIPRGISFS 2304
```

Search completed: December 22, 2003, 17:32:43
Job time : 20.9333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:41:00 ; Search time 32.4567 Seconds
(without alignments)
53.778 Million cell updates/sec

Title: US-09-909-164-12
Perfect score: 54
Sequence: 1 EEWVPGMDYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq 19Jun03:*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*

2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*

3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*

4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*

5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*

6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*

7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*

8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*

9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*

10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*

11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*

12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*

13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*

14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*

15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*

16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*

17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*

18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*

19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*

20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*

21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*

22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|----------------------------|
| 1 | 52 | 96.3 | 11 | 23 | ABB80524 Hepatitis C virus |
| 2 | 52 | 96.3 | 11 | 23 | ABB80528 Hepatitis C virus |
| 3 | 52 | 96.3 | 11 | 23 | ABB80529 Hepatitis C virus |
| 4 | 52 | 96.3 | 11 | 23 | ABB80561 Hepatitis C virus |
| 5 | 52 | 96.3 | 11 | 23 | ABB80562 Hepatitis C virus |
| 6 | 47 | 87.0 | 11 | 23 | ABB80538 Hepatitis C virus |
| 7 | 47 | 87.0 | 11 | 23 | ABB80542 Hepatitis C virus |
| 8 | 47 | 87.0 | 11 | 23 | ABB80543 Hepatitis C virus |
| 9 | 46 | 85.2 | 11 | 23 | ABB80521 Hepatitis C virus |

| | | | | | |
|----|----|------|----|----|----------------------------|
| 10 | 46 | 85.2 | 11 | 23 | ABB80522 Hepatitis C virus |
| 11 | 46 | 85.2 | 11 | 23 | ABB80525 Hepatitis C virus |
| 12 | 46 | 85.2 | 11 | 23 | ABB80526 Hepatitis C virus |
| 13 | 46 | 85.2 | 11 | 23 | ABB80547 Hepatitis C virus |
| 14 | 46 | 85.2 | 11 | 23 | ABB80548 Hepatitis C virus |
| 15 | 46 | 85.2 | 11 | 23 | ABB80551 Hepatitis C virus |
| 16 | 46 | 85.2 | 11 | 23 | ABB80556 Hepatitis C virus |
| 17 | 46 | 85.2 | 11 | 23 | ABB80557 Hepatitis C virus |
| 18 | 46 | 85.2 | 11 | 23 | ABB80559 Hepatitis C virus |
| 19 | 46 | 85.2 | 11 | 23 | ABB80563 Hepatitis C virus |
| 20 | 46 | 85.2 | 11 | 23 | ABB80564 Hepatitis C virus |
| 21 | 46 | 85.2 | 11 | 23 | ABB80565 Hepatitis C virus |
| 22 | 46 | 85.2 | 11 | 23 | ABB80566 Hepatitis C virus |
| 23 | 46 | 85.2 | 11 | 23 | ABB80567 Hepatitis C virus |
| 24 | 46 | 85.2 | 11 | 23 | ABB80568 Hepatitis C virus |
| 25 | 45 | 83.3 | 11 | 23 | ABB80523 Hepatitis C virus |
| 26 | 45 | 83.3 | 11 | 23 | ABB80527 Hepatitis C virus |
| 27 | 45 | 83.3 | 11 | 23 | ABB80558 Hepatitis C virus |
| 28 | 45 | 83.3 | 11 | 23 | ABB80560 Hepatitis C virus |
| 29 | 44 | 81.5 | 11 | 23 | ABB80533 Hepatitis C virus |
| 30 | 44 | 81.5 | 11 | 23 | ABB80534 Hepatitis C virus |
| 31 | 41 | 75.9 | 11 | 23 | ABB80535 Hepatitis C virus |
| 32 | 41 | 75.9 | 11 | 23 | ABB80536 Hepatitis C virus |
| 33 | 41 | 75.9 | 11 | 23 | ABB80539 Hepatitis C virus |
| 34 | 41 | 75.9 | 11 | 23 | ABB80540 Hepatitis C virus |
| 35 | 40 | 74.1 | 11 | 23 | ABB80537 Hepatitis C virus |
| 36 | 40 | 74.1 | 11 | 23 | ABB80541 Hepatitis C virus |
| 37 | 40 | 74.1 | 11 | 23 | ABB80544 Hepatitis C virus |
| 38 | 40 | 74.1 | 11 | 23 | ABB80545 Hepatitis C virus |
| 39 | 40 | 74.1 | 11 | 23 | ABB80549 Hepatitis C virus |
| 40 | 40 | 74.1 | 11 | 23 | ABB80552 Hepatitis C virus |
| 41 | 40 | 74.1 | 11 | 23 | ABB80553 Hepatitis C virus |
| 42 | 39 | 72.2 | 11 | 23 | ABB80546 Hepatitis C virus |
| 43 | 39 | 72.2 | 11 | 23 | ABB80550 Hepatitis C virus |
| 44 | 39 | 72.2 | 11 | 23 | ABB80554 Hepatitis C virus |
| 45 | 39 | 72.2 | 11 | 23 | ABB80555 Hepatitis C virus |

ALIGNMENTS

RESULT 1
ABB80524

ID ABB80524 standard; peptide; 11 AA.

AC ABB80524;

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #4.

KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide; virucide.

OS Synthetic.

FT Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"

FT Misc-difference 9 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

PN WO200208251-A2.

XX 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US23169.

XX

PR 21-JUL-2000; 2000US-220101P.
PA (CORV-) CORVAS INT INC.
XX
XX Lim-wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease
XX
XX Claim 17; Page 64; 69pp; English.
XX
CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;
Query Match 96.3%; Score 52; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EEVVPXGMDYS 11
Db 1 EEVVPXGMDYS 11
RESULT 2
ABB80528
ID ABB80528 standard; peptide; 11 AA.
AC ABB80528;
XX
XX 08-OCT-2002 (first entry)
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #8.
DE Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX Synthetic.
XX Key Location/Qualifiers
FH Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 8
FT Misc-difference 9 /note= "D-form residue"
FT Misc-difference 11 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX WO200208251-A2.
XX 31-JAN-2002.
XX 19-JUL-2001; 2001WO-US23169.
XX 21-JUL-2000; 2000US-220101P.
XX (CORV-) CORVAS INT INC.
XX Lim-wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease
XX
XX Claim 17; Page 64; 69pp; English.

PT Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease
XX
XX Claim 17; Page 64; 69pp; English.
XX
CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;
Query Match 96.3%; Score 52; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EEVVPXGMDYS 11
Db 1 EEVVPXGMDYS 11
RESULT 3
ABB80529
ID ABB80529 standard; peptide; 11 AA.
AC ABB80529;
XX
XX 08-OCT-2002 (first entry)
XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #9.
DE Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX Synthetic.
XX Key Location/Qualifiers
FH Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 8
FT Misc-difference 9 /note= "D-form residue"
FT Misc-difference 11 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX WO200208251-A2.
XX 31-JAN-2002.
XX 19-JUL-2001; 2001WO-US23169.
XX 21-JUL-2000; 2000US-220101P.
XX (CORV-) CORVAS INT INC.
XX Lim-wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease
XX
XX Claim 17; Page 64; 69pp; English.

CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;

Query Match 96.3%; Score 52; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVVVPXGMDYS 11
|||||
DB 1 EVVVPXGMDYS 11

RESULT 4
ABB80561
ID ABB80561 standard; peptide; 11 AA.

XX AC ABB80561;
XX DT 08-OCT-2002 (first entry)
XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #41.
XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX KM virucide.
XX OS Synthetic.

XX FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT Misc-difference 8 residue 7"
FT Modified-site 8 /note= "D-form residue"
FT Modified-site 8 /note= "Oxymethionine"
FT Modified-site 11 /note= "C-terminal amide"
XX

PN WO200208251-A2.

XX 31-JAN-2002.

PF 19-JUL-2001; 2001WO-US23169.

XX 21-JUL-2000; 2000US-220101P.

XX (CORV-) CORVAS INT INC.

XX PI Lim-wilby M, Levy OE, Brunck TK;

XX WPI; 2002-361643/39.

XX PT Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C
XX virus protease -

XX PS Claim 17; Page 65; 69pp; English.

XX CC The sequence represents a peptide compound of the invention having
XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX invention are alpha-ketoamide peptide analogues. The peptides have
XX virucide activity, and are useful for treating and in the manufacture of
XX a medicament to treat disorders associated with HCV protease. A
XX pharmaceutical composition comprising the peptide as an active ingredient

CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;

Query Match 96.3%; Score 52; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVVVPXGMDYS 11
|||||
DB 1 EVVVPXGMDYS 11

RESULT 5
ABB80562

ID ABB80562 standard; peptide; 11 AA.

XX AC ABB80562;

XX DT 08-OCT-2002 (first entry)

XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #42.

XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX KM virucide.

XX OS Synthetic.

XX FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT Misc-difference 8 residue 7"
FT Modified-site 8 /note= "D-form residue"
FT Modified-site 8 /note= "Oxymethionine"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX

PN WO200208251-A2.

XX 31-JAN-2002.

PF 19-JUL-2001; 2001WO-US23169.

XX 21-JUL-2000; 2000US-220101P.

XX (CORV-) CORVAS INT INC.

XX PI Lim-wilby M, Levy OE, Brunck TK;

XX WPI; 2002-361643/39.

XX PT Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C
XX virus protease -
XX PS Claim 17; Page 65; 69pp; English.
XX CC The sequence represents a peptide compound of the invention having
XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX invention are alpha-ketoamide peptide analogues. The peptides have
XX virucide activity, and are useful for treating and in the manufacture of
XX a medicament to treat disorders associated with HCV protease. A
XX pharmaceutical composition comprising the peptide as an active ingredient

Query Match 96.3%; Score 52; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EEVVPXGMDYS 11
 |||||
Db 1 EEVVPXGMDYS 11

RESULT 6
ABB80538
ID ABB80538 standard; peptide; 11 AA.
XX
AC ABB80538;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #18.
XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"

WO200208251-A2.
XX
PN 31-JAN-2002.
XX
PD 19-JUL-2001; 2001WO-US23169.
XX
PF 21-JUL-2000; 2000US-220101P.
XX
PR (CORV-) CORVAS INT INC.
XX
PI Lim-wilby M, Levy OE, Brunck TK;
XX
DR WPI; 2002-361643/39.
XX
CC Novel peptide compound having hepatitis C virus protease inhibitory activity useful for treating disorders associated with hepatitis C virus protease -
XX
PS Claim 17; Page 64; 69pp; English.
XX
CC The sequence represents a peptide compound of the invention having hepatitis C virus (HCV) protease inhibitory activity. The peptides of the invention are alpha-ketoamide peptide analogues. The peptides have virucide activity, and are useful for treating and in the manufacture of a medicament to treat disorders associated with HCV protease. A pharmaceutical composition comprising the peptide as an active ingredient is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;

Query Match 87.0%; Score 47; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.017;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMDYS 11
 |||||
Db 1 EEVVPXGMDYS 11

RESULT 7
ABB80542
ID ABB80542 standard; peptide; 11 AA.
XX
AC ABB80542;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #22.
XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Modified-site 11 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX
PN WO200208251-A2.
XX
PD 31-JAN-2002.
XX
PF 19-JUL-2001; 2001WO-US23169.
XX
PR 21-JUL-2000; 2000US-220101P.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Lim-wilby M, Levy OE, Brunck TK;
XX
DR WPI; 2002-361643/39.
XX
CC Novel peptide compound having hepatitis C virus protease inhibitory activity useful for treating disorders associated with hepatitis C virus protease -
XX
PS Claim 17; Page 65; 69pp; English.
XX
CC The sequence represents a peptide compound of the invention having hepatitis C virus (HCV) protease inhibitory activity. The peptides of the invention are alpha-ketoamide peptide analogues. The peptides have virucide activity, and are useful for treating and in the manufacture of a medicament to treat disorders associated with HCV protease. A pharmaceutical composition comprising the peptide as an active ingredient is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;
Query Match 87.0%; Score 47; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.017;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMDYS 11
 |||||
Db 1 EEVVPXGMDYS 11

RESULT 8
ABB80543
ID ABB80543 standard; peptide; 11 AA.
XX
AC ABB80543;
XX
DT 08-OCT-2002 (first entry)

FT Modified-site 11
XX /note= "C-terminal amide"
PN WO200208251-A2.
XX
PD 31-JAN-2002.
XX
PF 19-JUL-2001; 2001WO-US23169.
XX
PR 21-JUL-2000; 2000US-220101P.
XX
PA (CORV-) CORVAS INT INC.
XX
XX Lim-wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
DR
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
PS Claim 17; Page 64; 69pp; English.
XX
CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;
Query Match 85.2%; Score 46; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.026; 1; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGMDYS 11
DB ||||| ||
1 EEVVPXGMSYS 11
RESULT 11
ABB80525
ID ABB80525 standard; peptide; 11 AA.
XX
AC ABB80525;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #5.
XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX
OS Synthetic.
XX
PH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
FT
PN WO200208251-A2.
XX
PD 31-JAN-2002.
XX
PF 19-JUL-2001; 2001WO-US23169.
XX
PR 21-JUL-2000; 2000US-220101P.
XX
PA (CORV-) CORVAS INT INC.
XX
XX Lim-wilby M, Levy OE, Brunck TK;
PI 19-JUL-2001; 2001WO-US23169.

XX
PR 21-JUL-2000; 2000US-220101P.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Lim-wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
DR
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
PS Claim 17; Page 64; 69pp; English.
XX
CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;
Query Match 85.2%; Score 46; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.026; 1; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGMDYS 11
DB ||||| ||
1 EEVVPXGMSYS 11
RESULT 12
ABB80526
ID ABB80526 standard; peptide; 11 AA.
XX
AC ABB80526;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #6.
XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX
OS Synthetic.
XX
PH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
FT
PN WO200208251-A2.
XX
PD 31-JAN-2002.
XX
PF 19-JUL-2001; 2001WO-US23169.
XX
PR 21-JUL-2000; 2000US-220101P.
XX
PA (CORV-) CORVAS INT INC.
XX
XX Lim-wilby M, Levy OE, Brunck TK;
PI

XX DR WPI; 2002-361643/39.
 XX CC Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C
 PT virus protease -
 XX PS Claim 17; Page 64; 69pp; English.
 XX CC The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.
 XX PS Sequence 11 AA;
 XX SQ

Query Match 85.2%; Score 46; DB 23; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.026; 1; Indels 0; Gaps 0;
 Matches 10; Conservative 0; Mismatches 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 BEVVPXGMDYS 11
 DB 1 BEVVPXGMSYS 11
 |||||

RESULT 13
 ABB80547
 ID ABB80547 standard; peptide; 11 AA.
 AC ABB80547;
 XX 08-OCT-2002 (first entry)
 DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #27.
 XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.
 XX Synthetic.
 XX Key Location/Qualifiers
 FH Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Modified-site 11 /note= "C-terminal amide"
 FT WO200208251-A2.
 PN 31-JAN-2002.
 PD 19-JUL-2001; 2001WO-US23169.
 PF 21-JUL-2000; 2000US-220101P.
 PR (CORV-) CORVAS INT INC.
 XX Lim-wilby M, Levy OE, Brunck TK;
 XX WPI; 2002-361643/39.
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C
 PT virus protease -
 XX PS Claim 17; Page 65; 69pp; English.
 XX CC The sequence represents a peptide compound of the invention having

CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.
 XX PS Sequence 11 AA;
 XX SQ

Query Match 85.2%; Score 46; DB 23; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.026; 1; Indels 0; Gaps 0;
 Matches 10; Conservative 0; Mismatches 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 BEVVPXGMDYS 11
 DB 1 BEVVPXGTDYS 11
 |||||

RESULT 14
 ABB80548
 ID ABB80548 standard; peptide; 11 AA.
 AC ABB80548;
 XX 08-OCT-2002 (first entry)
 DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #28.
 XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.
 XX Synthetic.
 XX Key Location/Qualifiers
 FH Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Misc-difference 9 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"
 FT WO200208251-A2.
 PN 31-JAN-2002.
 PD 19-JUL-2001; 2001WO-US23169.
 PF 21-JUL-2000; 2000US-220101P.
 PR (CORV-) CORVAS INT INC.
 XX Lim-wilby M, Levy OE, Brunck TK;
 XX WPI; 2002-361643/39.
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C
 PT virus protease -
 XX PS Claim 17; Page 65; 69pp; English.
 XX CC The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.
 XX PS Sequence 11 AA;
 XX SQ

Query Match 85.2%; Score 46; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.026;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMDYS 11
| | | | | | | | | |
Db 1 EEVVPXGTDYS 11

RESULT 15
ABB80551
ID ABB80551 standard; peptide; 11 AA.
XX
AC ABB80551;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #31.
XX
DX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX
PN WO200208251-A2.
XX
PD 31-JAN-2002.
XX
PF 19-JUL-2001; 2001WO-US23169.
XX
PR 21-JUL-2000; 2000US-220101P.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Lim-wilby M, Levy OE, Brunck TK;
XX
DR WPI; 2002-361643/39.
XX
PT Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease
XX
PS Claim 17; Page 65; 69pp; English.
XX
CC The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;

Query Match 85.2%; Score 46; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.026;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMDYS 11
| | | | | | | | | |
Db 1 EEVVPXGSDYS 11

Search completed: December 22, 2003, 17:41:01
Job time : 32.4667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 17:24:36 ; Search time 9.06667 Seconds
(without alignments)
116.675 Million cell updates/sec

Title: US-09-909-164-12

Perfect score: 54

Sequence: 1 EVVFXGMDYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: PIR 76:**

1: pir1:**

2: pir2:**

3: pir3:**

4: pir4:**

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query | Length | DB | ID | Description |
|------------|-------|-------|--------|----|--------|--------------------|
| 1 | 40 | 74.1 | 156 | 2 | S54619 | hypothetical prote |
| 2 | 38 | 70.4 | 363 | 2 | D69551 | conserved hypothet |
| 3 | 36 | 66.7 | 102 | 2 | A2452 | V1 protein - tobac |
| 4 | 36 | 66.7 | 341 | 2 | S72649 | sucrose-phosphate |
| 5 | 36 | 66.7 | 348 | 2 | S72650 | sucrose-phosphate |
| 6 | 36 | 66.7 | 460 | 2 | G95764 | unknown protein P2 |
| 7 | 36 | 66.7 | 1049 | 2 | JC4783 | sucrose-phosphate |
| 8 | 36 | 66.7 | 1068 | 1 | J01329 | sucrose-phosphate |
| 9 | 36 | 66.7 | 1081 | 2 | T09837 | sucrose-phosphate |
| 10 | 36 | 66.7 | 1083 | 2 | T04062 | sucrose-phosphate |
| 11 | 36 | 66.7 | 1084 | 2 | T04103 | sucrose-phosphate |
| 12 | 35 | 64.8 | 425 | 2 | T24111 | hypothetical prote |
| 13 | 35 | 64.8 | 433 | 2 | H87660 | peptidoglycan-bind |
| 14 | 35 | 64.8 | 440 | 2 | H72784 | probable alkaline |
| 15 | 35 | 64.8 | 1150 | 2 | T20173 | hypothetical prote |
| 16 | 35 | 64.8 | 1474 | 2 | F69009 | probable membrane |
| 17 | 35 | 64.8 | 2747 | 2 | B49132 | fat facets (faf) s |
| 18 | 34 | 63.0 | 99 | 2 | S00210 | plastocyanin b - L |
| 19 | 34 | 63.0 | 155 | 2 | S38255 | plastocyanin b pre |
| 20 | 34 | 63.0 | 168 | 2 | S58208 | 6-O-methylguanine- |
| 21 | 34 | 63.0 | 290 | 2 | AG3104 | O6-methylguanine-D |
| 22 | 34 | 63.0 | 296 | 2 | F72745 | hypothetical prote |
| 23 | 34 | 63.0 | 357 | 1 | G69290 | probable hexosyltr |
| 24 | 34 | 63.0 | 366 | 2 | G69350 | L-lactate dehydrog |
| 25 | 34 | 63.0 | 366 | 2 | E86665 | ABC transporter AT |
| 26 | 34 | 63.0 | 565 | 2 | A70164 | phenylalanine-tRNA |
| 27 | 34 | 63.0 | 566 | 2 | F81138 | succinate dehydrog |
| 28 | 34 | 63.0 | 587 | 2 | F81138 | disease resistance |
| 29 | 34 | 63.0 | 906 | 2 | T48898 | disease resistance |

ALIGNMENTS

RESULT 1

S54619

hypothetical protein YOR013w - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein O2612; hypothetical protein YOL303.3

C:Species: Saccharomyces cerevisiae

C>Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002

C:Accession: S54619; S66879

R:de Haan, M.; Maarse, A.C.; Grivell, L.A.

submitted to the EMBL Data Library, May 1995

A:Reference number: S54617

A:Accession: S54619

A:Molecule type: DNA

A:Residues: 1-156 <DEH>

A:Cross-references: EMBL:X87331; NID:gl041652; PIDN:CAA60762.1; PID:g829123

R:de Haan, M.; Grivell, L.A.; Maarse, A.C.

submitted to the Protein Sequence Database, July 1996

A:Reference number: S66877

A:Accession: S66879

A:Molecule type: DNA

A:Residues: 1-156 <DEW>

A:Cross-references: EMBL:Z74920; NID:gl420109; PIDN:CAA99201.1; PID:gl420111; MIPS:YOF

A:Experimental source: strain S288C

C:Genetics:

A:Cross-references: SGD:S0005539

A:Map position: 15R

C:Superfamily: hypothetical protein YOR013w

| | | | | |
|-----------------------|-------|--------------|-------|-------------|
| Query Match | 74.1% | Score 40; | DB 2; | Length 156; |
| Best Local Similarity | 77.8% | Pred. No. 1; | | |
| Matches | 7; | Conservative | 1; | Mismatches |
| | | | | Indels |
| | | | | 0; |
| | | | | Gaps |
| | | | | 0; |

QY 2 EVVFXGMDY 10

DB 50 EVVFXGMDY 58

RESULT 2

D69551

conserved hypothetical protein AF2411 - Archaeoglobus fulgidus

C:Species: Archaeoglobus fulgidus

C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999

C:Accession: D69551

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.

Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Attiach, P.; Kaine, B.P.; Sykes,

Smith, H.O.; Woese, C.R.; Venter, J.C.

A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archa

A:Reference number: A69250; MUID:98049343; PMID:9389475

A:Accession: D69551

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA
A:Residues: 1-363 <KLE>
A:Cross-references: GB:AE001109; GB:AE000782; NID:G2689432; PIDN:AAB91255.1; PID:G265068

Query Match 70.4%; Score 38; DB 2; Length 363;
Best Local Similarity 54.5%; Pred. No. 6.8;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGMDYS 11
DB 120 ENIVPXGIDFS 130

RESULT 3
A42452
VI protein - tobacco yellow dwarf virus (strain Australia)
C:Species: tobacco yellow dwarf virus
C>Date: 15-Jan-1993 #sequence_revision 15-Jan-1993 #text_change 08-Oct-1999
C:Accession: A42452
R:Morris, B.A.M.; Richardson, K.A.; Haley, A.; Zhan, X.; Thomas, J.E.
Virology 187, 633-642, 1992
A:Title: The nucleotide sequence of the infectious cloned DNA component of tobacco yellow
A:Reference number: A42452; MUID:92188538; PMID:1546458
A:Accession: A42452
A:Molecule type: DNA
A:Residues: 1-102 <MOR>
A:Cross-references: GB:M81103; NID:G335283; PIDN:AAA47947.1; PID:G335284

Query Match 66.7%; Score 36; DB 2; Length 102;
Best Local Similarity 60.0%; Pred. No. 4.3;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVVPXGMDYS 11
DB 7 QVVPXGMDYS 16

RESULT 4
S72649
sucrose-phosphate synthase (EC 2.4.1.14) isoform 2 - Citrus unshiu (fragment)
C:Species: Citrus unshiu
C>Date: 24-Oct-1998 #sequence_revision 24-Oct-1998 #text_change 21-Jul-2000
C:Accession: S72649
R:Komatsu, A.; Takanokura, Y.; Omura, M.; Akihama, T.
Mol. Gen. Genet. 252, 346-351, 1996
A:Title: Cloning and molecular analysis of cDNAs encoding three sucrose phosphate synthase
A:Reference number: S72649; MUID:96439842; PMID:8842155
A:Accession: S72649
A:Molecule type: mRNA
A>Status: nucleic acid sequence not shown
A:Residues: 1-341 <KOM>
A:Cross-references: EMBL:AB006319; NID:G2588891; PIDN:BAA23215.1; PID:G2588892
A:Experimental source: fruit, cv. Miyagawa-Wase
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1997
C:Genetics:
A:Gene: SPS2
C:Function:
A:Description: catalyzes formation of sucrose-6-phosphate from UDPglucose and D-fructose
A:Pathway: sucrose biosynthesis
C:Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology
C:Keywords: glycosyltransferase; hexosyltransferase; sucrose biosynthesis
F:1-341/Domain: sucrose/sucrose-phosphate synthase homology (fragment) <SSPS>

Query Match 66.7%; Score 36; DB 2; Length 341;
Best Local Similarity 66.7%; Pred. No. 16;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
DB 228 VIPPGMDFS 236

RESULT 5
JC4783
sucrose-phosphate synthase (EC 2.4.1.14) - rice
C:Species: Oryza sativa (rice)

S72650
sucrose-phosphate synthase (EC 2.4.1.14) isoform 3 - Citrus unshiu (fragment)
C:Species: Citrus unshiu
C>Date: 24-Oct-1998 #sequence_revision 24-Oct-1998 #text_change 21-Jul-2000
C:Accession: S72650
R:Komatsu, A.; Takanokura, Y.; Omura, M.; Akihama, T.
Mol. Gen. Genet. 252, 346-351, 1996
A:Title: Cloning and molecular analysis of cDNAs encoding three sucrose phosphate synthase
A:Reference number: S72648; MUID:96439842; PMID:8842155
A:Accession: S72650
A>Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-348 <KOM>
A:Cross-references: EMBL:AB006660; NID:G2351059; PIDN:BAA22071.1; PID:G2351060
A:Experimental source: fruit, cv. Miyagawa-Wase
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1997
C:Genetics:
A:Gene: SPS3
C:Function:
A:Description: catalyzes formation of sucrose-6-phosphate from UDPglucose and D-fructose
A:Pathway: sucrose biosynthesis
C:Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology
C:Keywords: glycosyltransferase; hexosyltransferase; sucrose biosynthesis
F:1-348/Domain: sucrose/sucrose-phosphate synthase homology (fragment) <SSPS>

Query Match 66.7%; Score 36; DB 2; Length 348;
Best Local Similarity 66.7%; Pred. No. 17;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
DB 234 VIPPGMDFS 242

RESULT 6
G96764
unknown protein F25P22.17 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: G96764
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso
Chin, N.F.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzial
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: G96764
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-460 <STO>
A:Cross-references: GB:AB005173; NID:G6692750; PIDN:AAF24856.1; GSPDB:GN00141
C:Genetics:
A:Gene: F25P22.17
A:Map position: 1

Query Match 66.7%; Score 36; DB 2; Length 460;
Best Local Similarity 70.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMDY 10
DB 218 BEDVPSAMDY 227

C>Date: 10-May-1996 #sequence_revision 16-Aug-1996 #text_change 18-Jun-1999
 C:Accession: JCA4783
 R:Valdez-Alarcon, J.J.; Ferrando, M.; Salerno, G.; Jimenez-Moraila, B.; Herrera-Estrella
 Gene 170, 217-222, 1996
 A>Title: Characterization of a rice sucrose-phosphate synthase-encoding gene.
 A:Reference number: JCA4783; MUID:96235138; PMID:8666248

A:Accession: JCA4783
 A:Molecule type: mRNA
 A:Residues: 1-1049 <VAL>
 A:Cross-references: GB:U33175; NID:G1449931; PIDN:AAC49379.1; PID:G988270
 A:Note: UDPglucose-fructose-phosphate glucosyltransferase; Sucrosephosphate-UDPglucosylb
 C:Comment: This enzyme catalyzes the formation of sucrose-phosphate form UDP-glucose and
 C:Genetics:
 A:Gene: Sps1

C:Function: 24/1; 103/3; 183/3; 205/3; 435/3; 475/3; 519/3; 578/3; 596/3; 617/3; 931/3; 9
 A:Description: catalyzes the formation of sucrose-6-phosphate from UDPglucose and D-fruc
 A:Pathway: sucrose biosynthesis
 C:Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology
 C:Keywords: glycosyltransferase; hexosyltransferase; sucrose biosynthesis
 F:180-663/Domain: sucrose/sucrose-phosphate synthase homology <SSPS>

Query Match 66.7%; Score 36; DB 2; Length 1049;

Best Local Similarity 66.7%; Pred. No. 58;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11

DB 436 VIPFGMDFS 444

RESULT 8

JQ1329

sucrose-phosphate synthase (EC 2.4.1.14) - maize

C:Species: Zea mays (maize)

C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: JQ1329; PQ0260

R:Worrell, A.C.; Bruneau, J.M.; Summerfelt, K.; Boersig, M.; Voelker, T.A.

Plant Cell 3, 1121-1130, 1991

A>Title: Expression of a maize sucrose phosphate synthase in tomato alters leaf carbohyd
 A:Reference number: JQ1329; MUID:92338837; PMID:1840396

A:Accession: JQ1329

A:Molecule type: mRNA

A:Residues: 1-1068 <WOR>

A:Cross-references: GB:M97550; NID:G168625; PIDN:AAA33513.1; PID:G168626

A:Accession: PQ0260

A:Molecule type: protein

A:Residues: 71-74; 206-212; 471-481; 872-892 <WOR1>

C:Comment: This enzyme transfers the glucosyl group from UDPglucose to fructose-6-phosph
 C:Function: This enzyme is involved in the regulation of carbon partitioning in the leave

A:Description: catalyzes the formation of sucrose-6-phosphate from UDPglucose and D-fruc
 A:Pathway: sucrose biosynthesis

C:Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology

C:Keywords: glycosyltransferase; hexosyltransferase; sucrose biosynthesis

F:178-666/Domain: sucrose/sucrose-phosphate synthase homology <SSPS>

Query Match 66.7%; Score 36; DB 1; Length 1068;

Best Local Similarity 66.7%; Pred. No. 59;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11

DB 435 VIPFGMDFS 443

RESULT 9

TQ9837

sucrose-phosphate synthase (EC 2.4.1.14) isoform 2 - Craterostigma plantagineum

C:Species: Craterostigma plantagineum

C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000

C:Accession: TQ9837

R:Ingram, J.; Chandler, J.W.; Gallagher, L.; Salamini, F.; Bartels, D.

Plant Physiol. 115, 113-121, 1997

A>Title: Analysis of cDNA clones encoding sucrose-phosphate synthase in relation to su
 A:Reference number: 216874; MUID:97451773; PMID:9306694

A:Accession: TQ9837

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1081 <ING>

A:Cross-references: EMBL:Y11795; NID:G2190349; PIDN:CAA72491.1; PID:G2190350

A:Experimental source: ABA-treated callus

C:Genetics:

A:Gene: sps2

C:Function:

A:Description: catalyzes the formation of sucrose-6-phosphate from UDPglucose and D-fi
 A:Pathway: sucrose biosynthesis

C:Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology

C:Keywords: glycosyltransferase; hexosyltransferase; sucrose biosynthesis

F:176-674/Domain: sucrose/sucrose-phosphate synthase homology <SSPS>

Query Match 66.7%; Score 36; DB 2; Length 1081;

Best Local Similarity 66.7%; Pred. No. 60;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11

DB 445 VIPFGMDFS 453

RESULT 10

TQ4062

sucrose-phosphate synthase homolog F28M11.40 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 16-Jul-1999

C:Accession: TQ4062

R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.

submitted to the Protein Sequence Database, March 1999

A:Reference number: Z15184

A:Accession: TQ4062

A:Molecule type: DNA

A:Residues: 1-1083 <BEV>

A:Cross-references: EMBL:AL049487

A:Experimental source: cultivar Columbia; BAC clone F28M11

C:Genetics:

A:Map position: 4

A:Introns: 86/3; 116/3; 255/3; 322/2; 482/3; 526/3; 570/3; 629/3; 647/3; 668/3; 949/3;

A:Note: F28M11.40

C:Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology

F:230-714/Domain: sucrose/sucrose-phosphate synthase homology <SSPS>

Query Match 66.7%; Score 36; DB 2; Length 1083;

Best Local Similarity 66.7%; Pred. No. 60;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11

DB 483 VIPFGMDFS 491

RESULT 11

TQ4103

sucrose-phosphate synthase (EC 2.4.1.14) 1 - rice

C:Species: Oryza sativa (rice)

C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Jun-2000

C:Accession: TQ4103

R:Sakamoto, M.; Satozawa, T.; Kishimoto, N.; Higo, K.; Shimada, H.; Fujimura, T.

Plant Sci. 112, 207-217, 1995

A>Title: Structure and RFLP mapping of a rice sucrose phosphate synthase (SPS) gene th

A:Reference number: Z15212

A:Accession: TQ4103

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1084 <SAK>

A:Cross-references: EMBL:D45890; PIDN:BAA08304.1

A:Experimental source: subsp. Japonica

C;Genetics:
A;Gene: Sps1

A;Map position: 1
A;Introns: 120/3; 200/2; 221/3; 452/3; 492/3; 536/3; 595/3; 613/3; 634/3; 946/3; 989/2
C;Species: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology
C;Keywords: glycosyltransferase; hexosyltransferase
F;196-680/Domain: sucrose/sucrose-phosphate synthase homology <SSPS>

Query Match 66.7%; Score 36; DB 2; Length 1084;
Best Local Similarity 66.7%; Pred. No. 60;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMDYS 11
|:|:|:|:|
Db 453 VIPPGMDFS 461

RESULT 12

T24111

hypothetical protein R10D12.10 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T24111

R;Percy, C.

submitted to the EMBL Data Library, October 1996

A;Reference number: Z19842

A;Accession: T24111

A;Status: preliminary; translated from GB/EMBL/DBDJ

A;Molecule type: DNA

A;Residues: 1-425 <WIL>

A;Cross-references: EMBL:Z81109; PIDN:CAB03241.1; GSPDB:GN00023; CESP:R10D12.10

A;Experimental source: clone R10D12

C;Genetics:

A;Gene: CESP:R10D12.10

A;Map position: 5

A;Introns: 23/3; 56/3; 113/3; 257/2

Query Match 64.8%; Score 35; DB 2; Length 425;
Best Local Similarity 50.0%; Pred. No. 34;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 BEVVPXGMDY 10
|:|:|:|:|
Db 335 EQIVPGGLQY 344

RESULT 13

H87660

peptidoglycan-binding protein, probable [imported] - Caulobacter crescentus

C;Species: Caulobacter crescentus

C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001

C;Accession: H87660

R;Nieman, W.C.; Faldutylum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.

B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon

N. J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A;Title: Complete Genome Sequence of Caulobacter crescentus.

A;Reference number: H87249; MUID:21173698; PMID:11259647

A;Accession: H87660

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-433 <STO>

A;Cross-references: GB:AB005673; NID:gl3425020; PIDN:AAK25284.1; GSPDB:GN00148

C;Genetics:

A;Gene: CC3322

Query Match 64.8%; Score 35; DB 2; Length 433;
Best Local Similarity 54.5%; Pred. No. 34;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVVPXGMDYS 11
|:|:|:|:|
Db 266 EVILPFGFDYS 276

RESULT 14

H72784

probable alkaline proteinase APE0263 - Aeropyrum pernix (strain K1)

C;Species: Aeropyrum pernix

C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000

C;Accession: H72784

R;Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Tak

awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.;

DNA Res. 6, 83-101, 1999

A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aerop

A;Reference number: A72450; MUID:99310339; PMID:10382966

A;Accession: H72784

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-440 <KAW>

A;Cross-references: DDBJ:AF000058; NID:G5103388; PIDN:BAA79178.1; PID:G5103657

A;Experimental source: strain K1

C;Genetics:

A;Gene: APE0263

C;Superfamily: subtilisin; subtilisin homology

Query Match 64.8%; Score 35; DB 2; Length 440;
Best Local Similarity 66.7%; Pred. No. 35;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 EVVPXGMDY 10
|:|:|:|:|
Db 120 EVLPFGVDY 128

RESULT 15

T20173

hypothetical protein C53A5.2 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999

C;Accession: T20173; T23857

R;Mortimore, B.

submitted to the EMBL Data Library, November 1996

A;Reference number: Z19232

A;Accession: T20173

A;Status: preliminary; translated from GB/EMBL/DBDJ

A;Molecule type: DNA

A;Residues: 1-1150 <WIL>

A;Cross-references: EMBL:Z81486; PIDN:CAB03994.1; GSPDB:GN00023; CESP:C53A5.2

A;Experimental source: clone C53A5

R;Matthews, L.

submitted to the EMBL Data Library, August 1996

A;Reference number: Z19808

A;Accession: T23857

A;Status: preliminary; translated from GB/EMBL/DBDJ

A;Molecule type: DNA

A;Residues: 1-1150 <WIL>

A;Cross-references: EMBL:Z78015; PIDN:CAB01437.1; GSPDB:GN00023; CESP:C53A5.2

A;Experimental source: clone R02D5

C;Genetics:

A;Gene: CESP:C53A5.2

A;Map position: 5

A;Introns: 33/3; 63/3; 132/3; 169/3; 221/3; 299/3; 379/2; 423/2; 438/2; 471/1; 513/2;

Query Match

Best Local Similarity 64.8%; Score 35; DB 2; Length 1150;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMDYS 11
|:|:|:|:|
Db 562 VLPVGIDYS 570

Search completed: December 22, 2003, 17:44:57
Job time : 9.06667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:43:51 ; Search time 4.6 Seconds
(without alignments)
112.455 Million cell updates/sec

Title: US-09-909-164-12
Perfect score: 54
Sequence: 1 EVVVPXGMDYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--|
| 1 | 38 | 70.4 | 1058 | 1 | Q8986 fusobacteri |
| 2 | 36 | 66.7 | 102 | 1 | F1619 tobacco yel |
| 3 | 36 | 66.7 | 1049 | 1 | Q43802 oryza sativ |
| 4 | 36 | 66.7 | 1068 | 1 | P31927 zea mays (m |
| 5 | 36 | 66.7 | 1081 | 1 | Q04933 craterostig |
| 6 | 35 | 64.8 | 2778 | 1 | P55824 drosophila |
| 7 | 34.5 | 63.9 | 748 | 1 | Q9nr64 homo sapien |
| 8 | 34 | 63.0 | 154 | 1 | P20423 oryza sativ |
| 9 | 34 | 63.0 | 155 | 1 | P08248 hordeum vul |
| 10 | 34 | 63.0 | 168 | 1 | P11970 populus nig |
| 11 | 34 | 63.0 | 566 | 1 | P94283 borrelia bu |
| 12 | 34 | 63.0 | 908 | 1 | Q9fj88 arabidopsis |
| 13 | 34 | 63.0 | 908 | 1 | Q8w4j9 arabidopsis |
| 14 | 34 | 63.0 | 910 | 1 | P59584 arabidopsis |
| 15 | 33 | 61.1 | 276 | 1 | Q58349 methanococ |
| 16 | 33 | 61.1 | 283 | 1 | Q9hv69 pseudomonas |
| 17 | 33 | 61.1 | 394 | 1 | Q9kmv3 vibrio chol |
| 18 | 33 | 61.1 | 421 | 1 | P08503 rattus norv |
| 19 | 33 | 61.1 | 421 | 1 | Q52250 halomonas e |
| 20 | 33 | 61.1 | 423 | 1 | Q92507 halomonas e |
| 21 | 33 | 61.1 | 787 | 1 | O60344 homo sapien |
| 22 | 33 | 61.1 | 801 | 1 | Q61851 mus musculu |
| 23 | 33 | 61.1 | 806 | 1 | P18460 gallus gall |
| 24 | 33 | 61.1 | 877 | 1 | O74377 schizosacch |
| 25 | 33 | 61.1 | 982 | 1 | P27399 simian foam |
| 26 | 33 | 61.1 | 1401 | 1 | Q9kv29 vibrio chol |
| 27 | 33 | 61.1 | 2717 | 1 | P15822 homo sapien |
| 28 | 32.5 | 60.2 | 472 | 1 | ET2A XENLA |
| 29 | 32 | 59.3 | 97 | 1 | P20422 xenopus lae |
| 30 | 32 | 59.3 | 165 | 1 | TPX LISIN |
| 31 | 32 | 59.3 | 165 | 1 | Q92bc5 listeria in |
| 32 | 32 | 59.3 | 175 | 1 | Q8y6u8 listeria mo |
| 33 | 32 | 59.3 | 231 | 1 | PF02786; CPSase_L_D2; 2. PFam; PF02787; CPSase_L_D3; 1. |

RESULT 1

| ID | CARB_FUSNN | STANDARD; | PRT; | 1058 AA. |
|----|---|-----------|------|----------|
| AC | Q8RG86; | | | |
| DT | 28-FEB-2003 (Rel. 41, Created) | | | |
| DT | 28-FEB-2003 (Rel. 41, Last sequence update) | | | |
| DT | 28-FEB-2003 (Rel. 41, Last annotation update) | | | |
| DE | Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-phosphate synthetase ammonia chain). | | | |
| GN | CARB OR FN0422. | | | |
| OS | Fusobacterium nucleatum (subsp. nucleatum). | | | |
| OC | Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae; | | | |
| OC | Fusobacterium. | | | |
| OX | NCBI_TaxID=76856; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN=ATCC 25586; | | | |
| RX | MEDLINE=21886394; PubMed=11889109; | | | |
| RA | Kapatral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A., Bhatnagar A., Bartman A., Gardner W., Grechkin G., Zhu L., Vasileva O., Chu L., Kogan A., Chaga O., Goltzman E., Bernal A., Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R., Fonstein M., Kyripides N., Overbeek R., | | | |
| RT | "Genome sequence and analysis of the oral bacterium Fusobacterium nucleatum strain ATCC 25586."; | | | |
| RL | J. Bacteriol. 184:2005-2018(2002). | | | |
| CC | -/- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP + phosphate + L-glutamate + carbamoyl phosphate. | | | |
| CC | -/- COPACTOR: Binds 3 manganese ions per subunit (By similarity). | | | |
| CC | -/- PATHWAY: Arginine biosynthesis. | | | |
| CC | -/- SUBUNIT: Pyrimidine biosynthesis; first step. | | | |
| CC | -/- SUBUNIT: Composed of two chains; the small (or glutamine) chain promotes the hydrolysis of glutamine to ammonia, which is used by the large (or ammonia) chain to synthesize carbamoyl phosphate (By similarity). | | | |
| CC | -/- SIMILARITY: BELONGS TO THE CARB FAMILY. | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch). | | | |
| CC | EMBL; AB010554; AAL94625.1; ALT_INIT. | | | |
| DR | HMAP; MF 01210; -; 1. | | | |
| DR | InterPro; IPR006275; CarA_L_glu. | | | |
| DR | InterPro; IPR005483; CPhase_L. | | | |
| DR | InterPro; IPR005479; CPhase_L_D2. | | | |
| DR | InterPro; IPR005480; CPhase_L_D3. | | | |
| DR | InterPro; IPR005481; CPhase_L_N. | | | |
| DR | InterPro; IPR004362; MGS-like. | | | |
| DR | Pfam; PF02289; CPSase_L_chain; 2. | | | |
| DR | Pfam; PF02786; CPSase_L_D2; 2. | | | |
| DR | Pfam; PF02787; CPSase_L_D3; 1. | | | |

ALIGNMENTS

| | | | | | | |
|----|----|------|-----|---|------------|--------------------|
| 34 | 32 | 59.3 | 231 | 1 | ARAD_SALTY | P06190 salmonella |
| 35 | 32 | 59.3 | 233 | 1 | HIS9_THEMA | Q9wzr1 thermotoga |
| 36 | 32 | 59.3 | 288 | 1 | CGD2_RAT | Q04827 rattus norv |
| 37 | 32 | 59.3 | 289 | 1 | CGD2_HUMAN | P30279 homo sapien |
| 38 | 32 | 59.3 | 289 | 1 | CGD2_MOUSE | P30280 mus musculu |
| 39 | 32 | 59.3 | 291 | 1 | CGD1_BRARE | Q90459 brachydanio |
| 40 | 32 | 59.3 | 291 | 1 | CGD1_XENLA | P50755 xenopus lae |
| 41 | 32 | 59.3 | 291 | 1 | CGD2_CHICK | P49706 gallus gall |
| 42 | 32 | 59.3 | 291 | 1 | CGD2_XENLA | P53782 xenopus lae |
| 43 | 32 | 59.3 | 292 | 1 | CGD1_CHICK | P55169 gallus gall |
| 44 | 32 | 59.3 | 292 | 1 | CGD3_HUMAN | P30281 homo sapien |
| 45 | 32 | 59.3 | 295 | 1 | CGD1_HUMAN | P24385 homo sapien |

```

DR Pfam; PF02142; MGS; 1.
DR PRINTS; PR00098; CPSASE.
DR TIGRFAMs; TIGR01369; CPSaseII_lrg; 1.
DR PROSITE; PS00866; CPSASE_1; 2.
DR PROSITE; PS00867; CPSASE_2; 2.
KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
KW ATP-binding; Manganese; Complete proteome.
FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 402 546 OLIGOMERIZATION DOMAIN.
FT DOMAIN 547 929 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 930 1058 ALLOSTERIC DOMAIN.
FT REPEAT 1 546
FT REPEAT 547 1058
FT NP_BIND 153 210 ATP (POTENTIAL).
FT NP_BIND 302 352 ATP (POTENTIAL).
FT METAL 284 284 MANGANESE 1 (BY SIMILARITY).
FT METAL 298 298 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 300 300 MANGANESE 2 (BY SIMILARITY).
FT METAL 820 820 MANGANESE 3 (BY SIMILARITY).
FT METAL 832 832 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1058 AA; 117451 MW; ED7037AF77C1E39F CRC64;

Query Match 70.4%; Score 38; DB 1; Length 1058;
Best Local Similarity 60.0%; Pred. No. 9.1;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVPGXGMDYS 11
|:|:|:|:|:|
DB 190 EIVPGNLNYS 199

RESULT 2
Y1LK TYDVA STANDARD; PRT; 102 AA.
AC P31619;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-OCT-1993 (Rel. 27, Last annotation update)
DE Hypothetical 11.2 kDa protein.
GN VI.
OS Tobacco yellow dwarf virus (strain Australia) (TYDV).
OC Viruses; ssDNA viruses; Geminiviridae; Mastrevirus.
OX NCBI_TaxID=31599;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92188538; PubMed=1546458;
RA Morris B.A.M., Richardson K.A., Haley A., Zhan X., Thomas J.E.;
RT "The nucleotide sequence of the infectious cloned DNA component of tobacco yellow dwarf virus reveals features of geminiviruses infecting monocotyledonous plants.";
RL Virology 187:633-642(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M81103; AAA47947.1; -.
DR PIR; A2452; A2452.
DR InterPro; IPR002621; Gemini_mov.
DR Pfam; PF01708; Gemini_mov; 1.
KW Hypothetical protein.
SQ SEQUENCE 102 AA; 11178 MW; A40ECF1E0AF55B67 CRC64;

Query Match 66.7%; Score 36; DB 1; Length 102;
Best Local Similarity 60.0%; Pred. No. 2;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVPGXGMDYS 11
|:|:|:|:|:|

```

```

DB 7 QVPSGINYS 16

RESULT 3
SPS_ORYZA STANDARD; PRT; 1049 AA.
ID SPS_ORYZA
AC Q43802;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Sucrose-phosphate synthase (EC 2.4.1.14)
DE (UDP-glucose-fructose-phosphate glucosyltransferase).
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzoae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=cv. Indica-IR36; TISSUE=Leaf;
RX MEDLINE=96235138; PubMed=8666248;
RA Valdez-Alarcon J.J., Ferrando M., Jimenez-Moraila B.,
RA Herrera-Bstrrella L.;
RT "Characterization of a rice sucrose-phosphate synthase-encoding gene.";
RL Gene 170:217-222(1996).
CC -----
CC - FUNCTION: INVOLVED IN THE REGULATION OF CARBON PARTITIONING IN THE LEAVES OF PLANTS. MAY REGULATE THE SYNTHESIS OF SUCROSE AND THEREFORE PLAY A MAJOR ROLE AS A LIMITING FACTOR IN THE EXPORT OF PHOTOSYNTHETIC PRODUCTS OUT OF THE LEAF.
CC - CATALYTIC ACTIVITY: UDP-glucose + D-fructose 6-phosphate = UDP + sucrose 6-phosphate.
CC - ENZYME REGULATION: ACTIVITY REGULATED BY PHOSPHORYLATION AND MODERATED BY CONCENTRATION OF METABOLITES AND LIGHT.
CC - PATHWAY: Sucrose synthesis.
CC - SUBUNIT: Homodimer or homotetramer (By similarity).
CC - PTM: PHOSPHORYLATED. HOWEVER, PHOSPHORYLATION IS NOT ESSENTIAL FOR ENZYME FUNCTION (BY SIMILARITY).
CC - SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U33175; AAC49379.1; -.
DR PIR; JCA4783; JCA4783.
DR Grane; Q43802; -.
DR InterPro; IPR001296; Glyco_transf_1.
DR Pfam; PF00334; Glycosyltransferase; Phosphorylation.
KW Transferase; Glycosyltransferase; POLY-GLY.
FT DOMAIN 22 29 POLY-GLY.
FT DOMAIN 695 698 POLY-GLU.
FT DOMAIN 775 779 POLY-ARG.
SQ SEQUENCE 1049 AA; 116455 MW; ED862E2819AA4B04 CRC64;

Query Match 66.7%; Score 36; DB 1; Length 1049;
Best Local Similarity 66.7%; Pred. No. 24;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPGXGMDYS 11
|:|:|:|:|:|
DB 436 VIPPGMDFS 444

RESULT 4
SPS_MAIZE STANDARD; PRT; 1068 AA.
ID SPS_MAIZE
AC F31927;
DT 01-JUL-1993 (Rel. 26, Created)

```

DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 01-JUN-1994 (Rel. 29, Last annotation update)
 DE Sucrose-phosphate synthase (SC 2.4.1.14) (UDP-glucose-fructose-
 phosphate glucosyltransferase).
 GN SPS.
 OS Zea mays (Maize).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC PACCAD clade; Panicoideae; Andropogoneae; Zea.
 OX NCBI_TaxID=4577;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 71-74; 206-212; 471-481 AND
 RP 872-892.
 RC STRAIN=CV. PIONEER 3184; TISSUE=Leaf;
 RX MEDLINE=92338837; PubMed=1840396;
 RA Worrell A.C., Bruneau J.-M., Summerfelt K., Boerig M., Voelker T.A.;
 RT "Expression of a maize sucrose phosphate synthase in tomato alters
 RT leaf carbohydrate partitioning.";
 RL Plant Cell 3:1121-1130(1991).
 CC -!- FUNCTION: INVOLVED IN THE REGULATION OF CARBON PARTITIONING IN
 CC THE LEAVES OF PLANTS. MAY REGULATE THE SYNTHESIS OF SUCROSE AND
 CC THEREFORE PLAY A MAJOR ROLE AS A LIMITING FACTOR IN THE EXPORT OF
 CC PHOTOASSIMILATES OUT OF THE LEAF.
 CC -!- CATALYTIC ACTIVITY: UDP-glucose + D-fructose 6-phosphate = UDP +
 CC sucrose 6-phosphate.
 CC -!- ENZYME REGULATION: ACTIVITY REGULATED BY PHOSPHORYLATION AND
 CC MODERATED BY CONCENTRATION OF METABOLITES AND LIGHT.
 CC -!- PATHWAY: Sucrose synthesis.
 CC -!- SUBUNIT: HOMODIMER OR HOMOTETRAMER.
 CC -!- DEVELOPMENTAL STAGE: GERMINATING SEEDS OR MATURE LEAVES.
 CC -!- PTM: PHOSPHORYLATED. HOWEVER, PHOSPHORYLATION IS NOT ESSENTIAL FOR
 CC ENZYME FUNCTION.
 CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M97550; AAA33513.1; -.
 DR PIR; JQ1329; JQ1329.
 DR MaizeDB; 25294; -.
 DR InterPro; IPR001296; Glyco_transf_1.
 DR Pfam; PF00534; Glycosyltransferase; Phosphorylation.
 KW Transferase; Glycosyltransferase; Phosphorylation.
 FT DOMAIN 25 31 POLY-GLY.
 SQ SEQUENCE 1068 AA; 118575 MW; 07467985E9A1D282 CRC64;
 Query Match 66.7%; Score 36; DB 1; Length 1068;
 Best Local Similarity 66.7%; Pred. No. 24;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 3 VVPGMDYS 11
 Db 435 VVPGMDFS 443
 RESULT 5
 SPS2_CRAPL STANDARD; PRT; 1081 AA.
 ID SPS2_CRAPL
 AC 004933;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Sucrose-phosphate synthase 2 (SC 2.4.1.14) (UDP-glucose-fructose-
 phosphate glucosyltransferase 2).
 GN SPS2.
 OS Craterostigma plantagineum.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

OC Asteridae; lamids; Lamiales; Lamiales incertae sedis; Torenidae;
 OC Craterostigma.
 OX NCBI_TaxID=4153;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97451773; PubMed=9306694;
 RA Ingram J., Chandler J.W., Gallagher L., Salami F., Bartels D.;
 RT "Analysis of cDNA clones encoding sucrose-phosphate synthase in
 RT relation to sugar interconversions associated with dehydration in the
 RT resurrection plant Craterostigma plantagineum Hochst.";
 RL Plant Physiol. 115:113-121(1997).
 CC -!- FUNCTION: INVOLVED IN THE REGULATION OF CARBON PARTITIONING IN
 CC THE LEAVES OF PLANTS. MAY REGULATE THE SYNTHESIS OF SUCROSE AND
 CC THEREFORE PLAY A MAJOR ROLE AS A LIMITING FACTOR IN THE EXPORT OF
 CC PHOTOASSIMILATES OUT OF THE LEAF.
 CC -!- CATALYTIC ACTIVITY: UDP-glucose + D-fructose 6-phosphate = UDP +
 CC sucrose 6-phosphate.
 CC -!- ENZYME REGULATION: ACTIVITY REGULATED BY PHOSPHORYLATION AND
 CC MODERATED BY CONCENTRATION OF METABOLITES AND LIGHT.
 CC -!- PATHWAY: Sucrose synthesis.
 CC -!- SUBUNIT: Homodimer or homotetramer (by similarity).
 CC -!- PTM: PHOSPHORYLATED. HOWEVER, PHOSPHORYLATION IS NOT ESSENTIAL FOR
 CC ENZYME FUNCTION (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Y11795; CAA72491.1; -.
 DR PIR; T09837; T09837.
 DR InterPro; IPR001296; Glyco_transf_1.
 DR Pfam; PF00534; Glycosyltransferase; Phosphorylation; Multigene family.
 KW Transferase; Glycosyltransferase; Phosphorylation; Multigene family.
 FT DOMAIN 245 248 POLY-SER.
 FT DOMAIN 256 264 POLY-GLU.
 FT DOMAIN 787 790 POLY-ARG.
 SQ SEQUENCE 1081 AA; 120933 MW; DD142DC2F1A72900 CRC64;
 Query Match 66.7%; Score 36; DB 1; Length 1081;
 Best Local Similarity 66.7%; Pred. No. 24;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 3 VVPGMDYS 11
 Db 445 VVPGMDFS 453
 RESULT 6
 FAF_DROME STANDARD; PRT; 2778 AA.
 ID FAF_DROME
 AC P55824; Q9V9T6; Q9V027;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable ubiquitin carboxyl-terminal hydrolase FAF (EC 3.1.2.15)
 DE (Ubiquitin thiolesterase FAF) (Ubiquitin-specific processing protease
 DE FAF) (Deubiquitinating enzyme FAF) (Fat facets protein).
 GN FAF OR BCNA1.D22582 OR CG1945.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3), AND TISSUE SPECIFICITY.
 RC TISSUE=Eye imaginal disk;
 RX MEDLINE=93202020; PubMed=1295747;
 RA Fischer-Vize J.A., Rubin G.M., Lehmann R.;

RT "The fat facets gene is required for Drosophila eye and embryo
 development.";
 RL Development 116:985-1000(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=107311132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Anantides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.P.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.C., Mortman J.R., Fandell W.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Baileon R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swierkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN [3]
 RP REVISIONS, AND ALTERNATIVE SPLICING.
 RC STRAIN=Berkeley;
 RX MEDLINE=22426069; PubMed=12537572;
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a
 RT systematic review.";
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
 RN [4]
 RP SEQUENCE OF 1089-2778 FROM N.A. (ISOFORM 1).
 RC STRAIN=Berkeley;
 RX MEDLINE=20196012; PubMed=107311138;
 RA Rubin G.M., Hong L., Brokstein P., Evans-Holm M., Frise E.,
 RA Stapleton M., Harvey D.A.;
 RT "A Drosophila complementary DNA resource.";
 RL Science 287:2222-2224(2000).
 CC -1- FUNCTION: REQUIRED FOR EYE AND EMBRYO DEVELOPMENT, AND PLAYS A
 CC ROLE IN COMPOUND EYE ASSEMBLY AND OOGENESIS RESPECTIVELY. IN THE
 CC LARVAL EYE DISKS, CELLS OUTSIDE THE ASSEMBLING FACETS REQUIRE THIS
 CC PROTEIN FOR SHORT-RANGE CELL INTERACTIONS THAT PREVENT THE MYSTERY
 CC CELLS FROM BECOMING PHOTORECEPTORS. IT IS ALSO REQUIRED FOR
 CC NUCLEAR MIGRATION AND CELLULARIZATION IN EARLY EMBRYOGENESIS AND

CC COULD PLAY A ROLE IN POLE CELL DETERMINATION, DEVELOPMENT OR
 CC FUNCTION.
 CC -1- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
 CC ubiquitin + a thiol.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;
 CC Comment=Experimental confirmation may be lacking for some
 CC isoforms;
 CC Name=1;
 CC IsoId=P55824-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=P55824-2; Sequence=VSP_005270;
 CC Name=3;
 CC IsoId=P55824-3; Sequence=VSP_005269;
 CC -1- TISSUE SPECIFICITY: EYE DISKS AND OVARIES.
 CC -1- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically.
 CC -1- SIMILARITY: Belongs to peptidase family C19.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L04959; AAF01345.1; -;
 CC EMBL; L04958; AAF01346.1; -;
 CC EMBL; L04960; AAF01347.1; -;
 CC EMBL; L04960; AAF01348.1; -;
 CC EMBL; AE003779; AAF57198.1; -;
 CC EMBL; AE003779; AAF57198.1; -;
 CC EMBL; AF145677; AAN14291.1; -;
 CC MEROPS; C19.007; -;
 CC FlyBase; FBgn0005632; faf.
 CC GO; GO:0005737; C:cytoplasm; IDA.
 CC GO; GO:0007349; P:cellularization; IMP.
 CC GO; GO:0009795; P:embryonic morphogenesis; IMP.
 CC GO; GO:0008583; P:mystery cell fate differentiation (sensu Dr. .; IMP.
 CC GO; GO:0007097; P:nuclear migration; IMP.
 CC GO; GO:0006512; P:ubiquitin cycle; IGI.
 CC InterPro; IPR001394; UCH-2.
 CC Pfam; PF00443; UCH; 1.
 CC PROSITE; PS00972; UCH_2_1; 1.
 CC PROSITE; PS00973; UCH_2_2; 1.
 CC PROSITE; PS02335; UCH_2_3; 1.
 CC Ubl conjugation pathway; Hydrolase; Thiol protease;
 CC Developmental protein; Vision; Alternative splicing.
 CC By SIMILARITY.
 CC ACT_SITE 1677 1677
 CC ACT_SITE 1978 1978
 CC ACT_SITE 1986 1986
 CC VARSPLOC 2705 2778
 CC KRRVILKKLVESKDEEDATTATTAAATTEVTTSPTATATA
 CC ATLEPAGSELTWVEKNLIISOENPQAKSSIQ -> VTRA
 CC NNV (in isoform 3).
 CC /FTid=VSP_005269.
 CC IATAATLEPAGSELTWVEKNLIISOENPQAKSSIQ ->
 CC SORQOL (in isoform 2).
 CC /FTid=VSP_005270.
 CC E -> D (IN REF. 1).
 CC T -> S (IN REF. 1; AAF01345).
 CC FT CONFLICT 234 234
 CC FT CONFLICT 2725 2725
 CC FT CONFLICT 2778 2778
 CC FT SEQUENCE 2778 AA; 311139 MW; FFB90438BA53A02B CRC64;
 CC
 CC Query Match 64.8%; Score 35; DB 1; Length 2778;
 CC Best Local Similarity 54.5%; Pred No. 1.le-02;
 CC Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 CC
 CC QY 1 EVVVPKGMDS 11
 CC Db 1394 EVVDPGQDFS 1404
 CC
 CC RESULT 7
 CC KHL1_HUMAN

ID AC Q9NR64; Q9H4X4; Q9NR65; Q9P238; STANDARD; PRT; 748 AA.
DT 16-OCT-2001 (Rel. 40, Created)
DT 15-SEP-2003 (Rel. 40, Last sequence update)
DE Kelch-like protein 1.
GN KLHL1 OR KIAA1490.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20347694; PubMed=1088605;
RA Koob M.D., Nemes J.P., Benzow K.A.;
RT "The SCAB transcript is an antisense RNA to a brain-specific
transcript encoding a novel actin-binding protein (KLHL1).";
RL Hum. Mol. Genet. 9:1543-1551(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Brain;
RC MEDLINE=20277482; PubMed=10819331;
RA Nagase T., Kikuno R., Ishikawa K.-I., Hirosewa M., Ohara O.;
RT "Prediction of the coding regions of unidentified human genes. XVII.
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro.";
RL DNA Res. 7:143-150(2000).
RN [3]
RP SEQUENCE OF 179-409 FROM N.A.
RA Kay M.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAY PLAY A ROLE IN ORGANIZING THE ACTIN CYTOSKELETON OF
THE BRAIN CELLS.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: Highly expressed in brain.
CC -1- SIMILARITY: Contains 1 BTB/POZ domain.
CC -1- SIMILARITY: Contains 6 Kelch repeats.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF252283; AAF81719.1; -.
DR EMBL; AF252279; AAF81716.1; -.
DR EMBL; AB040923; BAA96014.1; ALT_INIT.
DR EMBL; AL353738; CAC16128.1; -.
DR GenBank; HGNC:6352; KLHL1.
DR MIM; 605332; -.
DR GO; GO:0003779; F-actin binding activity; NAS.
DR GO; GO:0030036; F-actin cytoskeleton organization and biogenesis; NAS.
DR InterPro; IPR000210; BTB_POZ.
DR InterPro; IPR006652; Kelch_rep.
DR Pfam; PF01344; Kelch; 6.
DR SMART; SM00225; BTB; 1.
DR SMART; SM00612; Kelch; 6.
DR PROSITE; PS00097; BTB; 1.
KW Cytoskeleton; Actin-binding; Kelch repeat; Repeat.
FT DOMAIN 43 88
FT DOMAIN 212 279
FT REPEAT 460 506
FT REPEAT 507 553
FT REPEAT 555 600
FT REPEAT 601 647
FT REPEAT 649 700
FT REPEAT 701 747
SQ SEQUENCE 748 AA; 82680 MW; C11C43D8282F9FF9 CRC64;
Query Match 63.9%; Score 34.5; DB 1; Length 748;
Best Local Similarity 80.0%; Pred. No. 34;
Matches 8; Conservative 1; Mismatches 0; Indels 1; Gaps 1;
Qy 1 EVVVPXGMDY 10
Db 127 EVVVP-GMDF 135
RESULT 8
PLAS ORYSA STANDARD; PRT; 154 AA.
ID AC P20423; Q9SBB8;
DT 01-FEB-1991 (Rel. 17, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Plastocyanin, chloroplast precursor.
GN PTEB.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzoideae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Ilpoom; TISSUE=Leaf;
RA Lee J.-S.;
RT "Molecular cloning and characterization of plastocyanin precursor in
rice";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 58-154.
RX STRAIN=cv. Japonica;
RC MEDLINE=89386623; PubMed=2780537;
RA Yano H., Kamo M., Tsugita A., Aso K., Nozu Y.;
RT "The amino acid sequence of plastocyanin from rice (Oryza sativa,
subspecies japonica).";
RL Protein Seq. Data Anal. 2:385-389(1989).
CC -1- FUNCTION: Participates in electron transfer between P700 and the
cytochrome b6-f complex in photosystem I.
CC -1- SUBCELLULAR LOCATION: LOOSELY BOUND TO THE INNER THYLAKOID
MEMBRANE SURFACE IN CHLOROPLASTS.
CC -1- SIMILARITY: Contains 1 plastocyanin-like domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF093636; AAC78108.1; -.
DR HSSP; P00289; 2PCF.
DR Gramene; P20423; -.
DR InterPro; IPR000923; BlueCu_1.
DR InterPro; IPR001235; Copper_blue.
DR Pfam; PF0127; copper-bind; 1.
DR PRINTS; PR00156; COPPERBLUE.
DR PRODOM; PD001235; Copper_blue; 1.
DR PROSITE; PS00196; COPPER_BLUE; 1.
KW Chloroplast; Electron transport; Copper; Thylakoid; Membrane;
FT TRANSIT 1 57
FT TRANSIT 58 154
FT CHAIN 58 154
FT DOMAIN 58 154
FT METAL 94 94
FT METAL 139 139
FT METAL 142 142
FT METAL 147 147
SQ SEQUENCE 154 AA; 15577 MW; E45725D5B5F400D CRC64;
Query Match 63.0%; Score 34; DB 1; Length 154;
Best Local Similarity 54.5%; Pred. No. 8.1;

Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMDYS 11
|: || |:
DB 100 EDAPVSGVDVS 110

RESULT 9

PLAS_HORVU STANDARD; PRT; 155 AA.

AC P08248;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Plastocyanin, chloroplast precursor.
GN PTBE.
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
OC Triticeae; Hordeum.
OX NCBI_TaxID=4513;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Bomi;
RA Nielsen O.S., Gausing K.;
RT "The precursor of barley plastocyanin: sequence of cDNA clones and
RT gene expression in different tissues.";
RL FEBS Lett. 225:159-162(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. NK 1558;
RX MEDLINE=94039081; PubMed=8223592;
RA Nielsen P., Gausing K.;
RT "In vitro binding of nuclear proteins to the barley plastocyanin gene
RT promoter region.";
RL Eur. J. Biochem. 217:97-104(1993).
CC -1- FUNCTION: Participates in electron transfer between P700 and the
CC cytochrome b6-f complex in photosystem I.
CC -1- SUBCELLULAR LOCATION: LOOSELY BOUND TO THE INNER THYLAKOID
CC MEMBRANE SURFACE IN CHLOROPLASTS.
CC -1- SIMILARITY: Contains 1 plastocyanin-like domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; Y00704; CAA68696.1; --
DR EMBL; Z28347; CAA82201.1; --
DR PIR; S38255; S38255.
DR HSP; P00289; 2PCF.
DR InterPro; IPR000923; BlueCu 1.
DR InterPro; IPR001235; Copper_blue.
DR Pfam; PF00127; copper-bind; 1.
DR PRINTS; PR00156; COPPERBLUE.
DR ProDom; PD001235; Copper_blue; 1.
DR PROSITE; PS00196; COPPER_BLUE; 1.
DR PROSITE; PS00196; COPPER_BLUE; 1.
KW Chloroplast; Electron transport; Copper; Thylakoid; Membrane;
KW Transit peptide.
FT TRANSIT 1 58 CHLOROPLAST.
FT CHAIN 59 155 PLASTOCYANIN.
FT DOMAIN 59 155 PLASTOCYANIN-LIKE.
FT METAL 95 95 COPPER (BY SIMILARITY).
FT METAL 140 140 COPPER (BY SIMILARITY).
FT METAL 143 143 COPPER (BY SIMILARITY).
FT METAL 148 148 COPPER (BY SIMILARITY).
FT VARIANT 120 120 T -> N (IN CV. NK 1558).
SQ SEQUENCE 155 AA; 15709 MW; DAA7EAB5F6F4F91 CRC64;

Query Match 63.0%; Score 34; DB 1; Length 155;

Best Local Similarity 54.5%; Pred. No. 8.2;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMDYS 11
|: || |:
DB 101 EDAPVSGVDVS 111

RESULT 10

PLAT_POPNI STANDARD; PRT; 168 AA.

AC P11970;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Plastocyanin B, chloroplast precursor.
GN PTBE.
OS Populus nigra (Lombardy poplar).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurocids I; Malpighiales; Salicaceae; Populus.
OX NCBI_TaxID=3691;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Italica; TISSUE=Leaf;
RA Reichert J., Jenzelewski V., Haehnel W.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 70-168.
RC STRAIN=cv. Italica;
RA Dimitrov M.I., Egorov C.A., Donchev A.A., Atanasov B.P.;
RT "Complete amino acid sequence of poplar plastocyanin B.";
RL FEBS Lett. 226:17-22(1987).
CC -1- FUNCTION: Participates in electron transfer between P700 and the
CC cytochrome b6-f complex in photosystem I.
CC -1- SUBCELLULAR LOCATION: LOOSELY BOUND TO THE INNER THYLAKOID
CC MEMBRANE SURFACE IN CHLOROPLASTS.
CC -1- SIMILARITY: THERE ARE 12 DIFFERENCES BETWEEN THE SEQUENCE OF
CC POPLAR PLASTOCYANINS A AND B.
CC -1- SIMILARITY: Contains 1 plastocyanin-like domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; Z50186; CAA90565.1; --
DR PIR; S00210; S00210.
DR PIR; S58208; S58208.
DR HSP; P00299; IPLC.
DR InterPro; IPR000923; BlueCu 1.
DR InterPro; IPR001235; Copper_blue.
DR Pfam; PF00127; copper-bind; 1.
DR PRINTS; PR00156; COPPERBLUE.
DR ProDom; PD001235; Copper_blue; 1.
DR PROSITE; PS00196; COPPER_BLUE; 1.
DR PROSITE; PS00196; COPPER_BLUE; 1.
KW Chloroplast; Electron transport; Copper; Thylakoid; Membrane;
KW Transit peptide; Multigene family.
FT TRANSIT 1 69 CHLOROPLAST.
FT CHAIN 70 168 PLASTOCYANIN B.
FT DOMAIN 70 168 PLASTOCYANIN-LIKE.
FT METAL 106 106 COPPER.
FT METAL 153 153 COPPER.
FT METAL 156 156 COPPER.
FT METAL 161 161 COPPER.
SQ SEQUENCE 168 AA; 16981 MW; F20DA6EA2038AEEA CRC64;

Query Match 63.0%; Score 34; DB 1; Length 168;
Best Local Similarity 54.5%; Pred. No. 8.9;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 EEVVPXGMDYS 11
|:|:|:|:
Db 112 EDAVPSGVDS 122

RESULT 11

ID SYFB BORBU STANDARD; PRT; 566 AA.
AC P94283;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phenylalanyl-tRNA synthetase beta chain (EC 6.1.1.20) (Phenylalanine--
DE tRNA ligase beta chain) (PheRS).
GN PHE1 OR BB0514.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RA Barbour A.G., Hinnebusch J.;
RT "Phenylalanyl-tRNA synthetase genes (alpha and beta subunits) and
RT thiorodoxin reductase gene of Borrelia burgdorferi."
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA Dougherty B., Tomb J.-P., Fleischmann R.D., Richardson D.,
RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA Van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA Utterback T., Wathley L., McDonald L., Artiach P., Bowman C.,
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT burgdorferi."
RL Nature 390:580-586 (1997).
CC -/- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA (Phe) = AMP +
CC diphosphate + L-phenylalanyl-tRNA (Phe).
CC -/- SUBUNIT: Tetramer of two alpha and two beta chains (By
CC similarity).
CC -/- SUBCELLULAR LOCATION: Cytoplasmic.
CC -/- SIMILARITY: BELONGS TO THE PHENYLALANYL-TRNA SYNTHETASE BETA
CC CHAIN FAMILY. SUBFAMILY 2.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U82978; AAB41019.1; -;
DR EMBL; AE001153; AAC66870.1; -;
DR PIR; A70164; A70164.
DR TIGR; BB0514; -;
DR HAMAP; MF_00284; -; 1.
DR InterPro; IPR005147; B5.
DR InterPro; IPR004531; PheT_arch.
DR Pfam; PF03484; B5; 1.
DR TIGRFAMs; TIGR00471; PheT_arch; 1.
KW Aminoacyl-tRNA synthetase, Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
SQ SEQUENCE 566 AA; 65173 MW; 9D48C8B5D6D3B74B CRC64;

Query Match 63.0%; Score 34; DB 1; Length 566;
Best Local Similarity 85.7%; Pred. No. 32;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 VPXGMDY 10
|:|:|:|:
Db 169 VPFGMDY 175

RESULT 12

ID BSL4 ARATH STANDARD; PRT; 908 AA.
AC Q9FJK8;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable disease resistance RPP8-like protein 4.
GN RPP8L4 OR AT5G48620 OR K15N18.9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Rukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=99087489; PubMed=9872454;
RA Nakamura Y., Sato S., Asamizu E., Kaneko T., Kotani H., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. VII.
RT Sequence features of the regions of 1,013,767 bp covered by sixteen
RT physically assigned P1 and TAC clones.";
RL DNA Res. 5:297-308 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,
RA Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,
RA Hayashizaki Y., Shinozaki K.;
RT "Arabidopsis thaliana full-length cDNA.";
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
CC -/- FUNCTION: Potential disease resistance protein.
CC -/- DOMAIN: The LRR repeats probably act as specificity determinant of
CC pathogen recognition (By similarity).
CC -/- SIMILARITY: Belongs to the disease resistance NB-LRR family.
CC RPP8/HRT subfamily.
CC -/- SIMILARITY: Contains 3 leucine-rich (LRR) repeats.
CC -/- SIMILARITY: Contains 1 NB-ARC domain.
CC -/- DATABASE: NAME-NIB-LRRS;
CC NOTE=Functional and comparative genomics of disease resistance gene
CC homologs;
CC WWW="http://niblrts.ucdavis.edu".

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AB015468; BAB10695.1; -;
DR EMBL; AK117163; BAC41841.1; -;
DR InterPro; IPR000767; Disease_resist.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR002182; NB-ARC.
DR Pfam; PF00560; LRR; 2.
DR Pfam; PF00931; NB-ARC; 1.
DR PRINTS; PR00364; DISEASERSIST.
KW Plant defense; ATP-binding; Repeat; Leucine-rich repeat.
FT DOMAIN 10 45
FT REPEAT 146 459 NB-ARC.
FT REPEAT 575 599 LRR 1.
FT REPEAT 600 623 LRR 2.
FT REPEAT 842 867 LRR 3.
FT NP_BIND 192 199 ATP (POTENTIAL).


```
FT VARIANT 27 31 IDGQL -> VDEQI (in cv. Landsberg erecta).
FT VARIANT 29 29 G -> E (in cv. Di-17).
FT VARIANT 85 87 SGK -> RGE (in cv. Di-17 and cv.
Landsberg erecta).
FT VARIANT 88 91 KGV -> EKG (in cv. Landsberg erecta).
FT VARIANT 100 100 C -> R (in cv. Landsberg erecta).
FT VARIANT 121 121 E -> D (in cv. Landsberg erecta).
FT VARIANT 129 129 F -> L (in cv. Di-17).
FT VARIANT 133 133 Q -> Q (in cv. Di-17).
FT VARIANT 138 138 G -> V (in cv. Landsberg erecta).
FT VARIANT 173 173 K -> T (in cv. Di-17).
FT VARIANT 177 177 G -> C (in cv. Di-17).
FT VARIANT 253 253 Y -> S (in cv. Di-17).
FT VARIANT 257 257 R -> P (in cv. Di-17).
FT VARIANT 265 265 A -> T (in cv. Di-17).
FT VARIANT 267 267 R -> K (in cv. Landsberg erecta).
FT VARIANT 270 270 V -> L (in cv. Di-17).
FT VARIANT 283 283 V -> R (in cv. Di-17).
FT VARIANT 369 369 A -> V (in cv. Di-17 and cv. Landsberg
erecta).
FT VARIANT 387 387 F -> S (in cv. Di-17 and cv. Landsberg
erecta).
FT VARIANT 399 399 W -> C (in cv. Di-17 and cv. Landsberg
erecta).
FT VARIANT 426 426 C -> R (in cv. Di-17 and cv. Landsberg
erecta).
FT VARIANT 429 429 N -> F (in cv. Di-17 and cv. Landsberg
erecta).
FT VARIANT 436 446 DSEISTYSLFY -> YSKISAYDLFN (in cv.
Landsberg erecta).
FT VARIANT 438 443 EISTYS -> KITQE (in cv. Di-17).
FT VARIANT 450 450 A -> V (in cv. Landsberg erecta).
FT VARIANT 460 460 E -> Q (in cv. Landsberg erecta).
FT VARIANT 480 486 DNYSMO -> NKYLRYH (in cv. Di-17).
FT VARIANT 480 481 DN -> NR (in cv. Landsberg erecta).
FT VARIANT 485 486 WQ -> SH (in cv. Landsberg erecta).
FT VARIANT 489 489 Y -> N (in cv. Landsberg erecta).
FT VARIANT 514 514 I -> K (in cv. Di-17 and cv. Landsberg
erecta).
FT VARIANT 518 518 C -> S (in cv. Di-17 and cv. Landsberg
erecta).
FT VARIANT 519 519 T -> I (in cv. Di-17).
FT VARIANT 528 528 S -> R (in cv. Di-17).
FT VARIANT 537 537 S -> R (in cv. Di-17).
FT VARIANT 542 542 H -> Q (in cv. Di-17).
FT VARIANT 543 543 I -> L (in cv. Landsberg erecta).
FT VARIANT 547 550 KNKT -> RNA (in cv. Di-17).
FT VARIANT 549 549 K -> N (in cv. Landsberg erecta).
FT VARIANT 558 565 PRFEEDYW -> WDEDFG (in cv. Landsberg
erecta).
FT VARIANT 558 560 PRF -> SRPK (in cv. Di-17).
FT VARIANT 564 564 Y -> F (in cv. Di-17).
FT VARIANT 584 584 S -> Y (in cv. Landsberg erecta).
FT VARIANT 595 595 C -> S (in cv. Di-17 and cv. Landsberg
erecta).
FT VARIANT 609 612 YEAK -> FLAG (in cv. Landsberg erecta).
FT VARIANT 630 635 NLRYDT -> DLSVHE (in cv. Di-17).
FT VARIANT 632 636 RVDPTE -> SVNKK (in cv. Landsberg erecta).
FT VARIANT 650 650 Q -> E (in cv. Di-17).
FT VARIANT 653 654 YL -> HI (in cv. Di-17).
FT VARIANT 659 659 Missing (in cv. Di-17).
FT VARIANT 660 660 M -> K (in cv. Landsberg erecta).
FT VARIANT 676 678 YLY -> FLF (in cv. Landsberg erecta).
FT VARIANT 678 679 YG -> FR (in cv. Di-17).
FT VARIANT 692 692 R -> H (in cv. Landsberg erecta).
FT VARIANT 697 697 R -> Q (in cv. Di-17).
FT VARIANT 700 700 A -> G (in cv. Di-17).

Query Match 63.0%; Score 34; DB 1; Length 908;
Best Local Similarity 60.0%; Pred. No. 53;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGMDY 10
```

```
Db 883 EKLVPQGEDY 892

RESULT 14
RPBH_ARATH
ID RPBH_ARATH STANDARD; PRT; 910 AA.
AC PS5584;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Disease resistance protein RPH8A (RPP8 homolog A).
GN RPH8A.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
SEQUENCE FROM N.A., AND FUNCTION.
RP STRAIN=cv. Landsberg erecta;
RC MEDLINE=99030193; PubMed=9811794;
RA McDowell J.M., Dhandaydham M., Long T.A., Aarts M.G.M., Goff S.,
RA Holub E.B., Dangl J.L.;
RT "Intragenic recombination and diversifying selection contribute to the
RT evolution of downy mildew resistance at the RPP8 locus of
RT Arabidopsis.";
RL Plant Cell 10:1861-1874(1998).
CC -I- FUNCTION: Disease resistance protein. Resistance proteins guard
CC the plant against pathogens that contain an appropriate avirulence
CC protein via an indirect interaction with this avirulence protein.
CC That triggers a defense system including the hypersensitive
CC response, which restricts the pathogen growth. In contrast to
CC RPP8, it does not specifically recognize the Emco5 avirulence
CC protein from Peronospora parasitica.
CC -I- MISCELLANEOUS: In cv. Columbia and cv. Di-17, this protein is not
CC present due to an unequal crossing over between the RPP8 and RPP8HA
CC genes that creates a unique RPP8 gene.
CC -I- SIMILARITY: Belongs to the disease resistance NB-LRR family.
CC RPP8/HRT subfamily.
CC -I- SIMILARITY: Contains 2 leucine-rich (LRR) repeats.
CC -I- SIMILARITY: Contains 1 NB-ARC domain.
CC -I- DATABASE: NAME=NIB-LRRS;
CC NOTE=Functional and comparative genomics of disease resistance gene
CC homologs;
CC WWW="http://niblr.ucdavis.edu".
KW Plant defense; ATP-binding; Repeat; Leucine-rich repeat.
FT DOMAIN 10 45 LEUCINE-ZIPPER.
FT DOMAIN 146 459 NB-ARC.
FT REPEAT 602 625 LRR 1.
FT REPEAT 844 869 LRR 2.
FT NP BIND 192 199 ATP (POTENTIAL).
SQ SEQUENCE 910 AA; 105263 MW; 5B1E9F65A19A12EE CRC64;

Query Match 63.0%; Score 34; DB 1; Length 910;
Best Local Similarity 60.0%; Pred. No. 53;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EEVVPXGMDY 10

Db 885 EKLVPQGEDY 894

RESULT 15
Y939_METJA
ID Y939_METJA STANDARD; PRT; 276 AA.
AC Q58349;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ0939.
GN MJ0939.
OS Methanococcus jannaschii.
```

```
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=868087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U67537; AAB98946.1; -.
DR PIR; C64417; C64417.
DR TIGR; MJ0939; -.
DR Pfam; PF03692; UPF0153; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 276 AA; 33454 MW; 97BD6D392BC8PDP CRC64;

Query Match          61.1%; Score 33; DB 1; Length 276;
Best Local Similarity 45.5%; Pred. No. 24;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Oy 1 BEVVPXGMDYS 11
   ||::||::|
Db 141 BEIIENGMEHS 151

Search completed: December 22, 2003, 17:42:26
Job time : 4.6 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 17:27:26 ; Search time 25.2 Seconds

(without alignments)

112.642 Million cell updates/sec

Title: US-09-909-164-12

Perfect score: 54

Sequence: 1 EEVVPXGMDYS 11

Storing table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 40 | 74.1 | 156 | 3 | Q12479 saccharomyc |
| 2 | 38 | 70.4 | 319 | 16 | Q8ESV7 |
| 3 | 38 | 70.4 | 363 | 17 | Q30260 archaeoglob |
| 4 | 38 | 70.4 | 1044 | 16 | Q8DIH0 |
| 5 | 36 | 66.7 | 341 | 10 | Q22081 citrus unsh |
| 6 | 36 | 66.7 | 348 | 10 | Q22096 citrus unsh |
| 7 | 36 | 66.7 | 452 | 10 | Q9C9T7 |
| 8 | 36 | 66.7 | 460 | 10 | Q9C9T7 |
| 9 | 36 | 66.7 | 1047 | 10 | P93782 saccharum o |
| 10 | 36 | 66.7 | 1083 | 10 | Q9SN30 |
| 11 | 36 | 66.7 | 1084 | 10 | Q43010 |
| 12 | 36 | 66.7 | 1100 | 10 | Q8S064 |
| 13 | 35 | 64.8 | 219 | 5 | Q9GQ04 |
| 14 | 35 | 64.8 | 253 | 16 | Q8XP88 |
| 15 | 35 | 64.8 | 298 | 2 | Q52367 |
| 16 | 35 | 64.8 | 368 | 2 | Q8KL43 |

| | | | | | |
|----|------|------|------|----|--------|
| 17 | 35 | 64.8 | 425 | 5 | Q9XVK4 |
| 18 | 35 | 64.8 | 433 | 16 | Q9A382 |
| 19 | 35 | 64.8 | 440 | 17 | Q9YFI3 |
| 20 | 35 | 64.8 | 511 | 2 | O52680 |
| 21 | 35 | 64.8 | 517 | 16 | Q8XZL5 |
| 22 | 35 | 64.8 | 595 | 10 | Q946J7 |
| 23 | 35 | 64.8 | 745 | 5 | Q95P46 |
| 24 | 35 | 64.8 | 1031 | 5 | Q9U6A3 |
| 25 | 35 | 64.8 | 1150 | 5 | O17704 |
| 26 | 35 | 64.8 | 1410 | 2 | O52673 |
| 27 | 35 | 64.8 | 1420 | 2 | O52666 |
| 28 | 35 | 64.8 | 1474 | 17 | O27146 |
| 29 | 35 | 64.8 | 1828 | 16 | Q98K29 |
| 30 | 34.5 | 63.9 | 748 | 4 | Q8TBJ7 |
| 31 | 34 | 63.0 | 143 | 5 | Q9VSV8 |
| 32 | 34 | 63.0 | 215 | 16 | Q8R9L5 |
| 33 | 34 | 63.0 | 222 | 16 | Q8RSL9 |
| 34 | 34 | 63.0 | 222 | 16 | Q8DZM9 |
| 35 | 34 | 63.0 | 284 | 2 | Q8GP33 |
| 36 | 34 | 63.0 | 290 | 16 | Q8U7J0 |
| 37 | 34 | 63.0 | 296 | 17 | Q9VET8 |
| 38 | 34 | 63.0 | 357 | 17 | O29920 |
| 39 | 34 | 63.0 | 366 | 17 | O29451 |
| 40 | 34 | 63.0 | 387 | 16 | Q98FX1 |
| 41 | 34 | 63.0 | 543 | 3 | Q8TFF4 |
| 42 | 34 | 63.0 | 558 | 16 | Q8R822 |
| 43 | 34 | 63.0 | 565 | 16 | Q9CINI |
| 44 | 34 | 63.0 | 587 | 16 | Q9JZP8 |
| 45 | 34 | 63.0 | 671 | 16 | Q8G3J2 |

ALIGNMENTS

RESULT 1

| | | | |
|--------|--|------|---------|
| Q12479 | PRELIMINARY; | PRT; | 156 AA. |
| ID | Q12479 | | |
| AC | Q12479; | | |
| DT | 01-NOV-1996 (TREMBLrel. 01, Created) | | |
| DT | 01-NOV-1996 (TREMBLrel. 01, Last sequence update) | | |
| DT | 01-JUN-2002 (TREMBLrel. 21, Last annotation update) | | |
| DE | ORF YOR013W. | | |
| GN | YOR013W. | | |
| OS | Saccharomyces cerevisiae (Baker's yeast). | | |
| OC | Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; | | |
| OC | Saccharomycetales; Saccharomycetaceae; Saccharomycetes. | | |
| OX | NCBI_TaxID=4932; | | |
| RN | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RA | De haan M., Grivell L.A., Maarse A.C.; | | |
| RL | Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases. | | |
| RN | [2] | | |
| RP | SEQUENCE FROM N.A. | | |
| RA | MIPS; | | |
| RL | Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases. | | |
| RN | [3] | | |
| RP | SEQUENCE FROM N.A. | | |
| RC | STRAIN=FY1679; | | |
| RA | De haan M., Maarse A.C., Grivell L.A.; | | |
| RL | Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases. | | |
| RN | [4] | | |
| RP | SEQUENCE FROM N.A. | | |
| RC | STRAIN=FY1679; | | |
| RX | MEDLINE=94019318; PubMed=8413243; | | |
| RA | Dumont M.E., Schlichter J.B., Cardillo T.S., Hayes M.K., Bethlenny G., | | |
| RA | Sherman P.; | | |
| RT | *CYC2 encodes a factor involved in mitochondrial import of yeast | | |
| RT | Cytochrome c.; | | |
| RL | Mol. Cell. Biol. 13:6442-6451 (1993). | | |
| RN | [5] | | |
| RP | SEQUENCE FROM N.A. | | |
| RC | STRAIN=FY1679; | | |
| RX | MEDLINE=94169519; PubMed=7764548; | | |


```

RA Lee Y.S., Shimizu J., Yoda K., Yamasaki M.;
RT "Molecular cloning of a gene, DHS1, which complements a drug-
RT hypersensitive mutation of the yeast Saccharomyces cerevisiae.";
RL Biosci. Biotechnol. Biochem. 58:391-395(1994).
DR EMBL; Z74920; CAA99201.1; -.
DR EMBL; X87331; CAA60762.1; -.
DR SGD; S0005539; YOR013W.
SQ SEQUENCE 156 AA; 17881 MW; 380442B74C272B41 CRC64;

Query Match 74.1%; Score 40; DB 3; Length 156;
Best Local Similarity 77.8%; Pred. No. 2.3;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVPXGMDY 10
DB 50 EVNPLGMDY 58
|||:|||||
|:|:|:|

RESULT 2
QBESV7 PRELIMINARY; PRT; 319 AA.
AC QBESV7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical conserved protein.
GN OB0509.
OS Oceanobacillus iheyensis.
OC Bacteria; Firmicutes; Bacillales; Oceanobacillus.
OX NCBI_TaxID=182710;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HTE831 / DSM 14371 / JCM 11309;
RX MEDLINE=22220767; PubMed=12335376;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya
RT Ridge and its unexpected adaptive capabilities to extreme
RT environments.";
RL Nucleic Acids Res. 30:3927-3935(2002).
DR EMBL; AP004594; BAC12465.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 319 AA; 35617 MW; 3BDAE4BF13879B37 CRC64;

Query Match 70.4%; Score 38; DB 16; Length 319;
Best Local Similarity 60.0%; Pred. No. 14;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 EVVPXGMDY 10
DB 189 EQLVPHGIDY 198
|||:|||||
|:|:|:|

RESULT 3
O30260 PRELIMINARY; PRT; 363 AA.
AC O30260;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein AF2411.
GN AF2411.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Winn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kurlavsky A.R., Graham D.E., Kyprides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,

```

```

RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeill L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.P., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
DR EMBL; AR001109; AAB91255.1; -.
DR TIGR; AF2411; -.
DR InterPro; IPR002103; Bac_luciferase.
DR Pfam; PF00296; bac_luciferase; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 363 AA; 41736 MW; 0B976EAE788F4803 CRC64;

Query Match 70.4%; Score 38; DB 17; Length 363;
Best Local Similarity 54.5%; Pred. No. 16;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMDYS 11
DB 120 ENIVEYGIDFS 130
|||:|||||
|:|:|:|

RESULT 4
QBDIH0 PRELIMINARY; PRT; 1044 AA.
ID QBDIH0
AC QBDIH0;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Multidrug efflux transporter.
GN TLL1618.
OS Synechococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=32046;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-1;
RX MEDLINE=22225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsumoto A., Nakazaki N.,
RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
RT Thermosynechococcus elongatus BP-1.";
RL DNA Res. 9:123-130(2002).
DR EMBL; AP005374; BAC09170.1; -.
KW Complete proteome.
SQ SEQUENCE 1044 AA; 113205 MW; 00B9C13F0F636D2P CRC64;

Query Match 70.4%; Score 38; DB 16; Length 1044;
Best Local Similarity 63.6%; Pred. No. 53;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMDYS 11
DB 843 BEVLPNGIGYS 853
|||:|||||
|:|:|:|

RESULT 5
O22081 PRELIMINARY; PRT; 341 AA.
ID O22081
AC O22081;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Sucrose-phosphate synthase (Fragment).
GN CITSPS2.
OS Citrus unshiu (Satsuma orange).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurooids II; Sapindales; Rutaceae; Citrus.

RN NCBI_TaxID=55188;

[1]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Miyagawa-Wase; TISSUE=Juice sacs and segment epidermis;

RX MEDLINE=96439842; PubMed=8842155;

RA Komatsu A., Takanokura Y., Omura M., Akihama T.;

RT "Cloning and molecular analysis of cDNA encoding three sucrose

phosphate synthase isoforms from a citrus fruit (Citrus unshiu

Marc.).";

RL Mol. Gen. Genet. 252:346-351(1996).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Miyagawa-Wase; TISSUE=Juice sacs and segment epidermis;

RA Komatsu A., Takanokura Y., Moriguchi T., Omura M., Akihama T.;

RT "Differential expression of three sucrose-phosphate synthase isoforms

during sucrose accumulation in citrus fruits (Citrus unshiu Marc.).";

RL Plant Sci. 140:169-178(1999).

DR EMBL; AB006319; BAA23215.1; -.

FT NON_TER 1 1

FT NON_TER 341 341

SQ SEQUENCE 341 AA; 38136 MW; 61417A69C4560777 CRC64;

Query Match 66.7%; Score 36; DB 10; Length 341;

Best Local Similarity 66.7%; Pred. No. 40;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11

Db

228 VPPGMDFS 236

RESULT 6

ID O22096 PRELIMINARY; PRT; 348 AA.

AC O22096;

DT 01-JAN-1998 (TREMBlrel. 05, Created)

DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE Sucrose-phosphate synthase (Fragment).

GN CitSps3.

OS Citrus unshiu (Satsuma orange).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurooids II; Sapindales; Rutaceae; Citrus.

OX NCBI_TaxID=55188;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Juice sacs and segment epidermis;

RX MEDLINE=96439842; PubMed=8842155;

RA Komatsu A., Takanokura Y., Omura M., Akihama T.;

RT "Cloning and molecular analysis of cDNA encoding three sucrose

phosphate synthase isoforms from a citrus fruit (Citrus unshiu

Marc.).";

RL Mol. Gen. Genet. 252:346-351(1996).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Juice sacs and segment epidermis;

RA Komatsu A., Takanokura Y., Moriguchi T., Omura M., Akihama T.;

RT "Differential expression of three sucrose-phosphate synthase isoforms

during sucrose accumulation in citrus fruits (Citrus unshiu Marc.).";

RL Plant Sci. 140:169-178(1999).

DR EMBL; AB006660; BAA22071.1; -.

FT NON_TER 1 1

FT NON_TER 348 348

SQ SEQUENCE 348 AA; 38556 MW; EE1C21BBA6FF5C5E CRC64;

Query Match 66.7%; Score 36; DB 10; Length 348;

Best Local Similarity 66.7%; Pred. No. 41;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11

Db 234 VPPGMDFS 242

[:| |||:]

RESULT 7

Q8W568 PRELIMINARY; PRT; 452 AA.

AC Q8W568;

DT 01-WAR-2002 (TREMBlrel. 20, Created)

DT 01-WAR-2002 (TREMBlrel. 20, Last sequence update)

DT 01-WAR-2003 (TREMBlrel. 23, Last annotation update)

DE Atig73750/F25P22.17.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurooids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.;

RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.;

RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.;

RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.;

RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.;

RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.;

RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.;

RA Ecker J.R.;

RT "Arabidopsis cDNA clones.;"

RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA Kim C.J., Chen H., Cheuk R., Shinn P., Banh J., Bowser L.;

RA Carninci P., Chang E., Dale J.M., Goldsmith A.D., Hayashizaki Y.;

RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.;

RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.;

RA Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.;

RA Tang C.C., Toriumi M., Wu H.C., Yamada K., Yamamura Y., Yu G., Yu S.;

RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;

RT "Arabidopsis ORF clones.;"

RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF419606; AAL31938.1; -.

DR EMBL; AY113044; AAM47352.1; -.

DR InterPro; IPR002471; Prol_endopep_ser.

DR InterPro; IPR000379; Ser_estr_sife.

DR PROSITE; PS00708; PRO_ENDOPEP_SER; 1.

SQ SEQUENCE 452 AA; 49682 MW; A159955B21742C4A CRC64;

Query Match 66.7%; Score 36; DB 10; Length 452;

Best Local Similarity 70.0%; Pred. No. 55;

Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGMDY 10

Db

210 BEDVPSNMDY 219

[:| |||:]

RESULT 8

Q9C9T7 PRELIMINARY; PRT; 460 AA.

ID Q9C9T7

DT 01-JUN-2001 (TREMBlrel. 17, Created)

DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)

DT 01-WAR-2003 (TREMBlrel. 23, Last annotation update)

DE Hypothetical 50.6 kDa protein.

GN F25P22.17.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurooids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

```

RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Egu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Uterback T., Van Aken S., Vaynsberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana.";
RL Nature 408:816-820(2000).
DR EMBL; AC012679; AGS52073.1; -.
DR InterPro; IPR002471; Prol_endopep_ser.
DR InterPro; PS000379; Ser_estra_site.
DR PROSITE; PS00708; PRO_ENDOPEP_SER; 1.
KW Hypothetical protein.
SQ SEQUENCE 460 AA; 50564 MW; E94B27B5C4B249EC CRC64;

Query Match 66.7%; Score 36; DB 10; Length 460;
Best Local Similarity 70.0%; Pred. No. 56;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVXPXGMDY 10
DB 218 EEDVFSAMDY 227

RESULT 9
P93782 PRELIMINARY; PRT; 1047 AA.
AC P93782;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Sucrose-phosphate synthase (EC 2.4.1.14) (Fragment).
GN SOSPS1.
OS Saccharum officinarum (Sugarcane).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC PACCAD clade; Panicoideae; Andropogoneae; Saccharum.
OX NCBI_TaxID=4547;
RN [1]_TaxID=4547;
RP SEQUENCE FROM N.A.
RC TISSUE=Leaf;
RA Sugiharto B., Sakakibara H., Sugiyama T.;
RT "Differential Expression of Two Genes for Sucrose-Phosphate Synthase
RT in Sugarcane: Molecular Cloning of the cDNAs and Comparative Analysis
RT of Gene Expression.";
RL Submitted (F8B-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB001337; BAA19241.1; -.
DR InterPro; IPR001296; Glyco_trans_1.
DR Pfam; PF00534; Glycos_transf_1; 1.
KW Glycosyltransferase; Transferase.
FT NON_TER 1
SQ SEQUENCE 1047 AA; 116379 MW; D0EDB34961ED83D CRC64;

Query Match 66.7%; Score 36; DB 10; Length 1047;
Best Local Similarity 66.7%; Pred. No. 1.4e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
DB 414 VIPPGMDFS 422

RESULT 10
Q9SN30 PRELIMINARY; PRT; 1083 AA.
AC Q9SN30;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Sucrose-phosphate synthase-like protein (EC 2.4.1.14).
GN F28W11.40 OR AT4G10120.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Murphy G., Ridley P., Hudson S., Bancroft I., Mewes H.W.,
RA Mayer K.F.X., Lemcke K., Schueller C.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL049487; CAB39764.1; -.
DR EMBL; AL161516; CAB78135.1; -.
DR InterPro; IPR001296; Glyco_trans_1.
DR Pfam; PF00534; Glycos_transf_1; 1.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 1083 AA; 122688 MW; EFC126FCA2137BB1 CRC64;

Query Match 66.7%; Score 36; DB 10; Length 1083;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
DB 483 VIPPGMDFS 491

RESULT 11
Q43010 PRELIMINARY; PRT; 1084 AA.
AC Q43010;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Sucrose phosphate synthase.
GN SPS1.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartioideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. japonica;
RA Sakamoto M., Satozawa T., Kishimoto N., Higo K., Shimada H.,
RA Fujimura T.;
RT "Structure and RFLP mapping of a rice sucrose phosphate synthase (SPS)
RT gene that is specifically expressed in the source organ.";
RL Plant Sci. 112:207-217(1995).
DR EMBL; D45890; BAA08304.1; -.
DR Gramene; Q43010; -.
DR InterPro; IPR001296; Glyco_trans_1.
DR InterPro; IPR001557; L_LDH-.
DR Pfam; PF00534; Glycos_transf_1; 1.
DR PROSITE; PS00064; L_LDH; 1.
SQ SEQUENCE 1084 AA; 119417 MW; 4D0D79AFEDD2F92B CRC64;

```

Query Match 66.7%; Score 36; DB 10; Length 1084;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
|:|:|:|:|
Db 453 VVPGMDFS 461

RESULT 12
Q8S064 PRELIMINARY; PRT; 1100 AA.
AC Q8S064;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative sucrose-phosphate synthase.
GN P06781.14.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzoae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
clone: P06781.1";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003437; BAB86107.1; -.
DR Gramene; Q8S064; -.
DR InterPro; IPR001296; Glyco_trans_1.
DR InterPro; IPR001557; L_LDH.
DR Pfam; PF00534; Glycos_transf_1; 1.
DR PROSITE; PS00064; L_LDH; 1.
SQ SEQUENCE 1100 AA; 121170 MW; 084F4604BA389CAD CRC64;

Query Match 66.7%; Score 36; DB 10; Length 1100;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
|:|:|:|:|
Db 469 VVPGMDFS 477

RESULT 13
Q9GQ04 PRELIMINARY; PRT; 219 AA.
AC Q9GQ04;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE Na+/K+/2Cl-cotransporter (Fragment).
OS Eriocheir sinensis (Chinese mitten crab).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC Brachyura; Grapsoidea; Varunidae; Eriocheir.
OX NCBI_TaxID=95602;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Gill;
RA Wehrauch D., Towle D.W.;
RT "Na+/H+-exchanger and Na+/K+/2Cl- -cotransporter are expressed in
gills of the euryhaline Chinese crab Eriocheir sinensis.";
RL Comp. Biochem. Physiol. 126:S158-S158 (2000).
DR EMBL; AF301160; AAG39938.1; -.
FT NON_TER 1
FT NON_TER 219
SQ SEQUENCE 219 AA; 24159 MW; 599442DA26FD3F1 CRC64;

Query Match 64.8%; Score 35; DB 5; Length 219;
Best Local Similarity 75.0%; Pred. NO. 39;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 VVPXGMDYS 11
|:|:|:|:|
Db 107 VPOGLDYS 114

RESULT 14
Q8XPA8 PRELIMINARY; PRT; 253 AA.
AC Q8XPA8;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein CPE0057.
GN CPE0057.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RX PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yanashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
flesh-eater";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001 (2002).
DR EMBL; AP003185; BAB79763.1; -.
DR InterPro; IPR000205; NAD binding.
DR InterPro; IPR000594; Thif_domain.
DR Pfam; PF00899; Thif; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 253 AA; 27714 MW; 814DF79D380D7486 CRC64;

Query Match 64.8%; Score 35; DB 16; Length 253;
Best Local Similarity 50.0%; Pred. NO. 46;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMDY 10
|:|:|:|:|
Db 108 BEIIPDDVDY 117

RESULT 15
O52367 PRELIMINARY; PRT; 298 AA.
AC O52367;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Aryl-alcohol dehydrogenase homolog (Fragment).
GN XYLBI.
OS Rhizobium tropici.
OG Plasmid pRtrCFN299a.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=398;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CFN299;
RA Rosenblueth M., Hynes M.F., Martinez-Romero E.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
CC -1- COPACTOR: ZINC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
FAMILY.
DR EMBL; AF036920; AAC04779.1; -.
DR InterPro; IPR002328; ADH_zinc.
DR InterPro; IPR002085; Adh_zn family.
DR InterPro; IPR000205; NAD_binding.

DR Pfam: PF00107; adh_zinc: 1.
DR PROSITE: PS00059; ADH_ZINC: 1.
KW Metal-binding; Oxidoreductase; Zinc; Plasmid.
FT NON_TER 298..298
SQ SEQUENCE 298 AA; 31092 MW; 49B2F8117C33AE87 CRC64;

Query Match 64.8%; Score 35; DB 2; Length 298;
Best Local Similarity 50.0%; Pred. No. 55;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 2 EVVPXGMDYS 11
|::|||:
Db 250 EIIPEGADFS 259

Search completed: December 22, 2003, 17:51:29
Job time : 25.2667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:41:00 ; Search time 32.4667 Seconds
(without alignments)
53.778 Million cell updates/sec

Title: US-09-909-164-13
Perfect score: 54
Sequence: 1 EEVVPXGMDYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues 1107863
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03.*

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|----------------------------|
| 1 | 52 | 96.3 | 11 | 23 | ABB80524 Hepatitis C virus |
| 2 | 52 | 96.3 | 11 | 23 | ABB80528 Hepatitis C virus |
| 3 | 52 | 96.3 | 11 | 23 | ABB80529 Hepatitis C virus |
| 4 | 52 | 96.3 | 11 | 23 | ABB80561 Hepatitis C virus |
| 5 | 52 | 96.3 | 11 | 23 | ABB80562 Hepatitis C virus |
| 6 | 47 | 87.0 | 11 | 23 | ABB80538 Hepatitis C virus |
| 7 | 47 | 87.0 | 11 | 23 | ABB80542 Hepatitis C virus |
| 8 | 47 | 87.0 | 11 | 23 | ABB80543 Hepatitis C virus |
| 9 | 46 | 85.2 | 11 | 23 | ABB80521 Hepatitis C virus |

| | | | | | |
|----|----|------|----|----|----------------------------|
| 10 | 46 | 85.2 | 11 | 23 | ABB80522 Hepatitis C virus |
| 11 | 46 | 85.2 | 11 | 23 | ABB80525 Hepatitis C virus |
| 12 | 46 | 85.2 | 11 | 23 | ABB80526 Hepatitis C virus |
| 13 | 46 | 85.2 | 11 | 23 | ABB80547 Hepatitis C virus |
| 14 | 46 | 85.2 | 11 | 23 | ABB80548 Hepatitis C virus |
| 15 | 46 | 85.2 | 11 | 23 | ABB80551 Hepatitis C virus |
| 16 | 46 | 85.2 | 11 | 23 | ABB80556 Hepatitis C virus |
| 17 | 46 | 85.2 | 11 | 23 | ABB80557 Hepatitis C virus |
| 18 | 46 | 85.2 | 11 | 23 | ABB80559 Hepatitis C virus |
| 19 | 46 | 85.2 | 11 | 23 | ABB80563 Hepatitis C virus |
| 20 | 46 | 85.2 | 11 | 23 | ABB80564 Hepatitis C virus |
| 21 | 46 | 85.2 | 11 | 23 | ABB80565 Hepatitis C virus |
| 22 | 46 | 85.2 | 11 | 23 | ABB80566 Hepatitis C virus |
| 23 | 46 | 85.2 | 11 | 23 | ABB80567 Hepatitis C virus |
| 24 | 46 | 85.2 | 11 | 23 | ABB80568 Hepatitis C virus |
| 25 | 45 | 83.3 | 11 | 23 | ABB80523 Hepatitis C virus |
| 26 | 45 | 83.3 | 11 | 23 | ABB80527 Hepatitis C virus |
| 27 | 45 | 83.3 | 11 | 23 | ABB80558 Hepatitis C virus |
| 28 | 45 | 83.3 | 11 | 23 | ABB80560 Hepatitis C virus |
| 29 | 44 | 81.5 | 11 | 23 | ABB80533 Hepatitis C virus |
| 30 | 44 | 81.5 | 11 | 23 | ABB80534 Hepatitis C virus |
| 31 | 41 | 75.9 | 11 | 23 | ABB80535 Hepatitis C virus |
| 32 | 41 | 75.9 | 11 | 23 | ABB80536 Hepatitis C virus |
| 33 | 41 | 75.9 | 11 | 23 | ABB80539 Hepatitis C virus |
| 34 | 41 | 75.9 | 11 | 23 | ABB80540 Hepatitis C virus |
| 35 | 40 | 74.1 | 11 | 23 | ABB80537 Hepatitis C virus |
| 36 | 40 | 74.1 | 11 | 23 | ABB80541 Hepatitis C virus |
| 37 | 40 | 74.1 | 11 | 23 | ABB80544 Hepatitis C virus |
| 38 | 40 | 74.1 | 11 | 23 | ABB80545 Hepatitis C virus |
| 39 | 40 | 74.1 | 11 | 23 | ABB80549 Hepatitis C virus |
| 40 | 40 | 74.1 | 11 | 23 | ABB80552 Hepatitis C virus |
| 41 | 40 | 74.1 | 11 | 23 | ABB80553 Hepatitis C virus |
| 42 | 39 | 72.2 | 11 | 23 | ABB80546 Hepatitis C virus |
| 43 | 39 | 72.2 | 11 | 23 | ABB80550 Hepatitis C virus |
| 44 | 39 | 72.2 | 11 | 23 | ABB80554 Hepatitis C virus |
| 45 | 39 | 72.2 | 11 | 23 | ABB80555 Hepatitis C virus |

ALIGNMENTS

RESULT 1
ABB80524

ID ABB80524 standard; peptide; 11 AA.

AC ABB80524;

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #4.

KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide; virucide.

OS Synthetic.

PH Key Location/Qualifiers

FT Modified-site 1

FT Modified-site 6 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"

FT Misc-difference 9

FT Modified-site 11 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

PN WO200208251-A2.

PD 31-JAN-2002.

PP 19-JUL-2001; 2001WO-US23169.

XX

PR 21-JUL-2000; 2000US-220101P.
 XX (CORV-) CORVAS INT INC.
 XX Lim-wilby M, Levy OE, Brunck TK;
 XX WPI; 2002-361643/39.
 XX
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C
 PT virus protease -
 XX
 PS Claim 17; Page 64; 69pp; English.
 XX
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.
 XX
 SQ Sequence 11 AA;
 Query Match 96.3%; Score 52; DB 23; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.0018;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 BEVVPXGMDYS 11
 Db 1 BEVVPXGMDYS 11
 RESULT 2
 ABB80528
 ID ABB80528 standard; peptide; 11 AA.
 XX
 AC ABB80528;
 XX
 XX 08-OCT-2002 (first entry)
 XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #8.
 DE Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 XX virucide.
 KW
 XX Synthetic.
 OS
 XX Key Location/Qualifiers
 FH Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Misc-difference 8 /note= "D-form residue"
 FT Misc-difference 9 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"
 FT
 FT WO200208251-A2.
 PN
 XX 31-JAN-2002.
 PD
 XX 19-JUL-2001; 2001WO-US23169.
 PF
 XX 21-JUL-2000; 2000US-220101P.
 PR
 XX (CORV-) CORVAS INT INC.
 PA
 XX Lim-wilby M, Levy OE, Brunck TK;
 XX WPI; 2002-361643/39.
 XX

PT Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C
 PT virus protease -
 XX
 PS Claim 17; Page 64; 69pp; English.
 XX
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.
 XX
 SQ Sequence 11 AA;
 Query Match 96.3%; Score 52; DB 23; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.0018;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 BEVVPXGMDYS 11
 Db 1 BEVVPXGMDYS 11
 RESULT 3
 ABB80529
 ID ABB80529 standard; peptide; 11 AA.
 XX
 AC ABB80529;
 XX
 XX 08-OCT-2002 (first entry)
 XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #9.
 DE Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 XX virucide.
 KW
 XX Synthetic.
 OS
 XX Key Location/Qualifiers
 FH Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Misc-difference 8 /note= "D-form residue"
 FT Misc-difference 9 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"
 FT
 FT WO200208251-A2.
 PN
 XX 31-JAN-2002.
 PD
 XX 19-JUL-2001; 2001WO-US23169.
 PF
 XX 21-JUL-2000; 2000US-220101P.
 PR
 XX (CORV-) CORVAS INT INC.
 PA
 XX Lim-wilby M, Levy OE, Brunck TK;
 XX WPI; 2002-361643/39.
 XX
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C
 PT virus protease -
 XX
 PS Claim 17; Page 64; 69pp; English.
 XX

CC The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.
 XX
 XX Sequence 11 AA;

Query Match 96.3%; Score 52; DB 23; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.0018;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVVVPXGMDYS 11
 DB 1 EVVVPXGMDYS 11

RESULT 4
 ABB80561
 ID ABB80561 standard; peptide; 11 AA.

XX AC ABB80561;
 XX DT 08-OCT-2002 (first entry)
 XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #41.
 XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 XX virucide.
 XX OS Synthetic.

XX FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Misc-difference 8 /note= "D-form residue"
 FT Modified-site 8 /note= "Oxymethionine"
 FT Modified-site 11 /note= "C-terminal amide"

XX PN WO200208251-A2.
 XX 31-JAN-2002.

XX PF 19-JUL-2001; 2001WO-US23169.
 XX PR 21-JUL-2000; 2000US-220101P.

XX PA (CORV-) CORVAS INT INC.

XX PI Lim-wilby M, Levy OE, Brunck TK;

XX DR WPI; 2002-361643/39.

XX PT Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C
 PT virus protease -

XX PS Claim 17; Page 65; 69pp; English.

XX CC The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient

CC is useful for treating disorders associated with hepatitis C virus.
 XX
 SQ Sequence 11 AA;

Query Match 96.3%; Score 52; DB 23; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.0018;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVVVPXGMDYS 11
 DB 1 EVVVPXGMDYS 11

RESULT 5
 ABB80562
 ID ABB80562 standard; peptide; 11 AA.

XX AC ABB80562;
 XX DT 08-OCT-2002 (first entry)
 XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #42.
 XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 XX virucide.
 XX OS Synthetic.

XX FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Misc-difference 8 /note= "D-form residue"
 FT Modified-site 8 /note= "Oxymethionine"
 FT Misc-difference 9 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"

XX PN WO200208251-A2.

XX PD 31-JAN-2002.

XX PF 19-JUL-2001; 2001WO-US23169.

XX PR 21-JUL-2000; 2000US-220101P.

XX PA (CORV-) CORVAS INT INC.

XX PI Lim-wilby M, Levy OE, Brunck TK;

XX DR WPI; 2002-361643/39.

XX PT Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C
 PT virus protease -

XX PS Claim 17; Page 65; 69pp; English.

XX CC The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.

XX SQ Sequence 11 AA;

Query Match 96.3%; Score 52; DB 23; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EEVVPXGMDYS 11
|||||
Db 1 EEVVPXGMDYS 11

RESULT 6

ABB80538
ID ABB80538 standard; peptide; 11 AA.
XX
AC ABB80538;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #18.
XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "N-terminal acetyl"
FT Misc-difference 9
FT Modified-site 11 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
FT
FT
FT
PN WO200208251-A2.
XX
XX 31-JAN-2002.
XX
PP 19-JUL-2001; 2001WO-US23169.
XX
PR 21-JUL-2000; 2000US-220101P.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Lim-wilby M, Levy OE, Brunck TK;
XX
DR WPI; 2002-361643/39.
XX
PT Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
PS Claim 17; Page 64; 69pp; English.
XX

Sequence 11 AA;

The sequence represents a peptide compound of the invention having hepatitis C virus (HCV) protease inhibitory activity. The peptides of the invention are alpha-ketoamide peptide analogues. The peptides have virucide activity, and are useful for treating and in the manufacture of a medicament to treat disorders associated with HCV protease. A pharmaceutical composition comprising the peptide as an active ingredient is useful for treating disorders associated with hepatitis C virus.

Query Match 87.0%; Score 47; DB 23; Length 11;

Best Local Similarity 90.9%; Pred. No. 0.017;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EEVVPXGMDYS 11
|||||
Db 1 EEVVPXGMDYS 11

RESULT 8

ABB80543
ID ABB80543 standard; peptide; 11 AA.
XX
AC ABB80543;
XX
DT 08-OCT-2002 (first entry)

RESULT 7

ABB80542
ID ABB80542 standard; peptide; 11 AA.
XX
AC ABB80542;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #22.
XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
FT
FT
FT
PN WO200208251-A2.
XX
XX 31-JAN-2002.
XX
PP 19-JUL-2001; 2001WO-US23169.
XX
PR 21-JUL-2000; 2000US-220101P.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Lim-wilby M, Levy OE, Brunck TK;
XX
DR WPI; 2002-361643/39.
XX
PT Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
PS Claim 17; Page 65; 69pp; English.
XX

Sequence 11 AA;

Query Match 87.0%; Score 47; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.017;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 EEVVPXGMDYS 11
|||||
Db 1 EEVVPXGMDYS 11

XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #23.
XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX KW virucide.
XX OS Synthetic.
XX FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6
FT Modified-site 11 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX WO200208251-A2.
XX 31-JAN-2002.
XX 19-JUL-2001; 2001WO-US23169.
XX 21-JUL-2000; 2000US-220101P.
XX (CORV-) CORVAS INT INC.
XX Lim-wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX Claim 17; Page 65; 69pp; English.
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX SQ Sequence 11 AA;
XX Query Match 87.0%; Score 47; DB 23; Length 11;
XX Best Local Similarity 90.9%; Pred. No. 0.017;
XX Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX 1 EEVVPXGMDYS 11
XX |||||
XX 1 EEVVPXGQDYS 11
XX
XX RESULT 9
XX ABB80521
XX ID ABB80521 standard; peptide; 11 AA.
XX AC ABB80521;
XX DT 08-OCT-2002 (first entry)
XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #1.
XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX KW virucide.

OS Synthetic.
XX FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Modified-site 11 /note= "C-terminal amide"
XX WO200208251-A2.
XX 31-JAN-2002.
XX 19-JUL-2001; 2001WO-US23169.
XX 21-JUL-2000; 2000US-220101P.
XX (CORV-) CORVAS INT INC.
XX Lim-wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX Claim 17; Page 64; 69pp; English.
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX SQ Sequence 11 AA;
XX Query Match 85.2%; Score 46; DB 23; Length 11;
XX Best Local Similarity 90.9%; Pred. No. 0.026;
XX Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX 1 EEVVPXGMDYS 11
XX |||||
XX 1 EEVVPXGMSYS 11
XX
XX RESULT 10
XX ABB80522
XX ID ABB80522 standard; peptide; 11 AA.
XX AC ABB80522;
XX DT 08-OCT-2002 (first entry)
XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #2.
XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX KW virucide.
XX OS Synthetic.
XX FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 9 /note= "D-form residue"
FT

FT Modified-site 11
FT /note= "C-terminal amide"
PN WO200208251-A2.
XX
XX
PD 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US23169.
XX
XX 21-JUL-2000; 2000US-220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
FT activity useful for treating disorders associated with hepatitis C
FT virus protease -
XX
XX Claim 17; Page 64; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;
Query Match 85.2%; Score 46; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.026; 1; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGMDYS 11
Db 1 EEVVPXGMSYS 11
RESULT 11
ABB80525
ID ABB80525 standard; peptide; 11 AA.
XX
XX ABB80525;
AC
XX 08-OCT-2002 (first entry)
DT
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #5.
XX
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX
XX Synthetic.
OS
XX Key Location/Qualifiers
FH Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
FT
FT WO200208251-A2.
PN
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US23169.
XX
XX 21-JUL-2000; 2000US-220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-wilby M, Levy OE, Brunck TK;
XX
XX 19-JUL-2001; 2001WO-US23169.

XX
XX 21-JUL-2000; 2000US-220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
FT activity useful for treating disorders associated with hepatitis C
FT virus protease -
XX
XX Claim 17; Page 64; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;
Query Match 85.2%; Score 46; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.026; 1; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGMDYS 11
Db 1 EEVVPXGMSYS 11
RESULT 12
ABB80526
ID ABB80526 standard; peptide; 11 AA.
XX
XX ABB80526;
AC
XX 08-OCT-2002 (first entry)
DT
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #6.
XX
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX
XX Synthetic.
OS
XX Key Location/Qualifiers
FH Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
FT
FT WO200208251-A2.
PN
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US23169.
XX
XX 21-JUL-2000; 2000US-220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-wilby M, Levy OE, Brunck TK;
XX
XX 19-JUL-2001; 2001WO-US23169.

XX DR WPI; 2002-361643/39.
 XX PT Novel peptide compound having hepatitis C virus protease inhibitory
 XX PT activity useful for treating disorders associated with hepatitis C
 XX PT virus protease -
 XX PS Claim 17; Page 64; 69pp; English.
 XX CC The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.
 XX SQ Sequence 11 AA;
 Query Match 85.2%; Score 46; DB 23; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.026;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 EEVVPXGMDYS 11
 DB ||||| |||
 1 EEVVPXGMDYS 11
 RESULT 13
 ABB80547
 ID ABB80547 standard; peptide; 11 AA.
 AC ABB80547;
 DT 08-OCT-2002 (first entry)
 DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #27.
 XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.
 XX Synthetic.
 XX Key Location/Qualifiers
 FH Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Modified-site 11 /note= "C-terminal amide"
 PN WO200208251-A2.
 XX 31-JAN-2002.
 XX 19-JUL-2001; 2001WO-US23169.
 XX 21-JUL-2000; 2000US-220101P.
 XX (CORV-) CORVAS INT INC.
 PI Lim-wilby M, Levy OE, Brunck TK;
 XX WPI; 2002-361643/39.
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 XX activity useful for treating disorders associated with hepatitis C
 XX virus protease -
 XX PS Claim 17; Page 65; 69pp; English.
 XX CC The sequence represents a peptide compound of the invention having

CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.
 XX SQ Sequence 11 AA;
 Query Match 85.2%; Score 46; DB 23; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.026;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 EEVVPXGMDYS 11
 DB ||||| |||
 1 EEVVPXGMDYS 11

RESULT 14
 ABB80548
 ID ABB80548 standard; peptide; 11 AA.
 AC ABB80548;
 DT 08-OCT-2002 (first entry)
 DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #28.
 XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.
 XX Synthetic.
 XX Key Location/Qualifiers
 FH Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Misc-difference 9 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"
 PN WO200208251-A2.
 XX 31-JAN-2002.
 XX 19-JUL-2001; 2001WO-US23169.
 XX 21-JUL-2000; 2000US-220101P.
 XX (CORV-) CORVAS INT INC.
 PI Lim-wilby M, Levy OE, Brunck TK;
 XX WPI; 2002-361643/39.
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 XX activity useful for treating disorders associated with hepatitis C
 XX virus protease -
 XX PS Claim 17; Page 65; 69pp; English.
 XX CC The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.
 XX SQ Sequence 11 AA;

Query Match 85.2%; Score 46; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.026; 1; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMDYS 11
Db 1 EEVVPXGTDYS 11

RESULT 15

ABB0551
ID ABB0551 standard; peptide; 11 AA.

AC ABB0551;

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #31.

KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
virucide.

OS Synthetic.

FH Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6

FT /note= "Norvalyl carbonyl forming keto-amide linkage with
residue 7"

FT Misc-difference 9

FT /note= "D-form residue"

FT Modified-site 11

FT /note= "C-terminal amide"

PN WO200208251-A2.

PD 31-JAN-2002.

PF 19-JUL-2001; 2001WO-US23169.

PR 21-JUL-2000; 2000US-220101P.

PA (CORV-) CORVAS INT INC.

PI Lim-wilby M, Levy OE, Brunck TK;

DR WPI; 2002-361643/39.

PT Novel peptide compound having hepatitis C virus protease inhibitory
activity useful for treating disorders associated with hepatitis C
virus protease

PS Claim 17; Page 65; 69pp; English.

CC The sequence represents a peptide compound of the invention having
hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
invention are alpha-ketoamide peptide analogues. The peptides have
virucide activity, and are useful for treating and in the manufacture of
a medicament to treat disorders associated with HCV protease. A
pharmaceutical composition comprising the peptide as an active ingredient
is useful for treating disorders associated with hepatitis C virus.

SQ Sequence 11 AA;

Query Match 85.2%; Score 46; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.026; 1; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EEVVPXGMDYS 11
Db 1 EEVVPXGTDYS 11

Search completed: December 22, 2003, 17:41:01
Job time : 32.4667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:37:03 ; Search time 10.1333 Seconds
(without alignments)
45.930 Million cell updates/sec

Title: US-09-909-164-13
Perfect score: 54
Sequence: 1 EEVVPXGMDYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCBUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|----------------------|
| 1 | 36 | 66.7 | 341 | 3 | US-08-853-948B-4 |
| 2 | 36 | 66.7 | 348 | 3 | US-08-853-948B-5 |
| 3 | 36 | 66.7 | 368 | 4 | US-09-697-367-24 |
| 4 | 36 | 66.7 | 1049 | 4 | US-09-394-272-10 |
| 5 | 36 | 66.7 | 1068 | 2 | US-08-429-054A-11 |
| 6 | 36 | 66.7 | 1068 | 2 | US-08-718-777-7 |
| 7 | 36 | 66.7 | 1068 | 3 | US-09-051-341-7 |
| 8 | 36 | 66.7 | 1068 | 4 | US-09-394-272-8 |
| 9 | 36 | 66.7 | 1081 | 4 | US-09-394-272-4 |
| 10 | 36 | 66.7 | 1083 | 4 | US-09-394-272-11 |
| 11 | 36 | 66.7 | 1084 | 4 | US-09-394-272-9 |
| 12 | 34 | 63.0 | 140 | 3 | US-08-569-147-76 |
| 13 | 34 | 63.0 | 140 | 3 | US-08-569-147-82 |
| 14 | 34 | 63.0 | 1065 | 4 | US-09-252-991A-31637 |
| 15 | 33 | 61.1 | 59 | 4 | US-08-963-851-14 |
| 16 | 33 | 61.1 | 378 | 1 | US-08-070-165F-8 |
| 17 | 33 | 61.1 | 378 | 2 | US-08-885-418-8 |
| 18 | 33 | 61.1 | 473 | 4 | US-09-252-991A-26805 |
| 19 | 33 | 61.1 | 765 | 4 | US-09-819-989-4 |
| 20 | 33 | 61.1 | 801 | 3 | US-09-383-630-6 |
| 21 | 33 | 61.1 | 811 | 4 | US-09-819-989-2 |
| 22 | 33 | 61.1 | 883 | 4 | US-09-667-373-4 |
| 23 | 32 | 59.3 | 65 | 6 | 5177197-51 |
| 24 | 32 | 59.3 | 102 | 2 | US-08-580-988A-23 |
| 25 | 32 | 59.3 | 152 | 2 | US-08-460-694-4 |
| 26 | 32 | 59.3 | 152 | 3 | US-08-460-744-4 |
| 27 | 32 | 59.3 | 152 | 3 | US-07-667-711B-4 |

28 59.3 173 1 US-08-193-977-7 Sequence 7, Appli
29 59.3 189 2 US-08-464-517-21 Sequence 21, Appl
30 59.3 189 2 US-08-246-361A-21 Sequence 21, Appl
31 59.3 189 3 US-08-463-772-21 Sequence 21, Appl
32 59.3 189 5 PCT-US93-05000-21 Sequence 21, Appl
33 59.3 231 3 US-08-926-842B-20 Sequence 20, Appl
34 59.3 236 2 US-08-464-517-22 Sequence 22, Appl
35 59.3 236 2 US-08-246-361A-22 Sequence 22, Appl
36 59.3 236 3 US-08-463-772-22 Sequence 22, Appl
37 59.3 236 5 PCT-US93-05000-22 Sequence 22, Appl
38 59.3 240 3 US-08-926-842B-21 Sequence 21, Appl
39 59.3 241 4 US-09-107-532A-4172 Sequence 4172, Ap
40 59.3 280 2 US-08-464-517-6 Sequence 6, Appli
41 59.3 280 3 US-08-463-772-6 Sequence 6, Appli
42 59.3 289 2 US-08-246-361A-4 Sequence 4, Appli
43 59.3 289 5 PCT-US93-05000-4 Sequence 4, Appli
44 59.3 291 5 PCT-US93-05000-6 Sequence 6, Appli
45 59.3 292 2 US-08-464-517-23 Sequence 23, Appl

ALIGNMENTS

RESULT 1
US-08-853-948B-4
; Sequence 4, Application US/08853948B
; Patent No. 6210943
; GENERAL INFORMATION:
; APPLICANT: AKIHAMA, Toyota
; TITLE OF INVENTION: SUCROSE PHOSPHATE SYNTHASE FROM CITRUS AND DNA ENCODING
; FILE REFERENCE: 0049-0235-0
; CURRENT APPLICATION NUMBER: US/08/853,948B
; CURRENT FILING DATE: 1997-05-09
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Citrus unshiu
; FEATURES:
; OTHER INFORMATION: Xaa at position 109 is one of Ala, Arg, Asn, Asp,
; OTHER INFORMATION: Cys, Gln, Glu, Gly, His, Ile, Leu, Lys, Met, Phe,
; OTHER INFORMATION: Pro, Ser, Thr, Tyr, or Val
US-08-853-948B-4

Query Match 66.7%; Score 36; DB 3; Length 341;
Best Local Similarity 66.7%; Pred. No. 18;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMDYS 11
Db 228 VPPGMDFS 236

RESULT 2
US-08-853-948B-5
; Sequence 5, Application US/08853948B
; Patent No. 6210943
; GENERAL INFORMATION:
; APPLICANT: AKIHAMA, Toyota
; TITLE OF INVENTION: SUCROSE PHOSPHATE SYNTHASE FROM CITRUS AND DNA ENCODING
; FILE REFERENCE: 0049-0235-0
; CURRENT APPLICATION NUMBER: US/08/853,948B
; CURRENT FILING DATE: 1997-05-09
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 348
; TYPE: PRT
; ORGANISM: Citrus unshiu
US-08-853-948B-5

Query Match 66.7%; Score 36; DB 3; Length 348;
Best Local Similarity 66.7%; Pred. No. 19;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
|:|:|:|:
DB 234 VPPGMDFS 242

RESULT 3

US-09-697-367-24
; Sequence 24, Application US/09697367
; Patent No. 6323015
; GENERAL INFORMATION:
; APPLICANT: Orozco Jr., Emil M.
; APPLICANT: Calmi, Perry G.
; APPLICANT: Weng, Zude
; APPLICANT: Tarczynski, Mitchell
; TITLE OF INVENTION: SUCROSE PHOSPHATE SYNTHASE
; FILE REFERENCE: BE1166 US NA
; CURRENT APPLICATION NUMBER: US/09/697,367
; CURRENT FILING DATE: 2000-10-26
; PRIOR APPLICATION NUMBER: 60/084,529
; PRIOR FILING DATE: 1998-MAY-07
; PRIOR APPLICATION NUMBER: PCT/US99/09865
; PRIOR FILING DATE: 1999-MAY-06
; NUMBER OF SEQ ID NOS: 24
; SEQ ID NO 24
; LENGTH: 368
; TYPE: PRT
; ORGANISM: Zea mays
US-09-697-367-24

Query Match 66.7%; Score 36; DB 4; Length 368;
Best Local Similarity 66.7%; Pred. No. 20;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
|:|:|:|:
DB 217 VPPGMDFS 225

RESULT 4

US-09-394-272-10
; Sequence 10, Application US/09394272
; Patent No. 6472588
; GENERAL INFORMATION:
; APPLICANT: Haigler, Candace H.
; APPLICANT: Holaday, A. Scott
; TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
; FILE REFERENCE: 201304/1000
; CURRENT APPLICATION NUMBER: US/09/394,272
; CURRENT FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1049
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-394-272-10

Query Match 66.7%; Score 36; DB 4; Length 1049;
Best Local Similarity 66.7%; Pred. No. 67;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
|:|:|:|:
DB 436 VPPGMDFS 444

RESULT 5

US-08-429-054A-11
; Sequence 11, Application US/08429054A
; Patent No. 5917126
; GENERAL INFORMATION:
; APPLICANT: VAN ASSCHE, CHARLES; LANDO, DANIELLE; BRUNEAU,
; APPLICANT: JEAN; VOELKER, TONI; GERVAIS, MONICA
; TITLE OF INVENTION: SUCROSE PHOSPHATE SYNTHETASE (SPS).
; TITLE OF INVENTION: A PREPARATION METHOD AND CDNA THEREFOR, AND USE OF THE
; TITLE OF INVENTION: CDNA FOR MODIFYING SPS EXPRESSION IN PLANT CELLS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIERMAN AND MUSERLIAN
; STREET: 600 THIRD AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/429,054A
; FILING DATE: 26-APR-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 842,337
; FILING DATE: 20-March-1992
; APPLICATION NUMBER: PCT/FR 91/00593
; FILING DATE: 18-July-1991
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: French 90402094.9
; FILING DATE: 20-July-1990
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Charles A. Musierlian
; REGISTRATION NUMBER: 19,683
; REFERENCE/DOCKET NUMBER: 146.1137
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 661-8000
; TELEFAX: (212) 661-8002
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1068
; TYPE: Amino acid
; STRANDEDNESS: Single
; TOPOLOGY: Unknown
; MOLECULE TYPE: Peptide
US-08-429-054A-11

Query Match 66.7%; Score 36; DB 2; Length 1068;
Best Local Similarity 66.7%; Pred. No. 68;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
|:|:|:|:
DB 435 VPPGMDFS 443

RESULT 6

US-08-718-777-7
; Sequence 7, Application US/08718777
; Patent No. 5981852
; GENERAL INFORMATION:
; APPLICANT: Van Assche, C.
; APPLICANT: Lando, D.
; APPLICANT: Bruneau, J. M.
; APPLICANT: Voelker, T.
; APPLICANT: Gervais, M.
; TITLE OF INVENTION: MODIFICATION OF SUCROSE

/ TITLE OF INVENTION: PHOSPHATE
/ TITLE OF INVENTION: SYNTHASE IN PLANTS
/ NUMBER OF SEQUENCES: 14
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Law Offices of Barbara Rae-Venter
/ STREET: 260 Sheridan Avenue, Suite 440
/ CITY: Palo Alto
/ STATE: California
/ COUNTRY: USA
/ ZIP: 94306
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/718,777
/ FILING DATE: NOT YET ASSIGNED
/ CLASSIFICATION: 800
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/175,471
/ FILING DATE: 27-DEC-1993
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Barbara Rae-Venter
/ REGISTRATION NUMBER: 32,750
/ REFERENCE/DOCKET NUMBER: CGNE.072.02US
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (415)328-4400
/ TELEFAX: (415)328-4477
/ INFORMATION FOR SEQ ID NO: 7:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1068 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-718-777-7

Query Match 66.7%; Score 36; DB 2; Length 1068;
Best Local Similarity 66.7%; Pred. No. 68;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMDYS 11
|:|:|:|:
Db 435 VPPGMDFS 443

RESULT 7
US-09-051-341-7
/ Sequence 7, Application US/09051341
/ Patent No. 6124528
/ GENERAL INFORMATION:
/ APPLICANT: Shewmaker, C. K.
/ TITLE OF INVENTION: MODIFICATION OF SOLUBLE SOLIDS USING
/ TITLE OF INVENTION: SUCROSE PHOSPHATE SYNTHASE ENCODING SEQUENCE
/ NUMBER OF SEQUENCES: 14
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Rae-Venter Law Group, P.C.
/ STREET: 260 Sheridan Avenue, Suite 440
/ CITY: Palo Alto
/ STATE: California
/ COUNTRY: USA
/ ZIP: 94306
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/051,341
/ FILING DATE:
/ CLASSIFICATION: 800
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: PCT/US96/17351

/ FILING DATE: 25-OCT-1996
/ APPLICATION NUMBER: US 08/549,016
/ FILING DATE: 27-OCT-1995
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/372,200
/ FILING DATE: 12-JAN-1995
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Barbara Rae-Venter, Ph.D.,
/ REGISTRATION NUMBER: 32,750
/ REFERENCE/DOCKET NUMBER: CGNE.110.02US
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (415)328-4400
/ TELEFAX: (415)328-4477
/ INFORMATION FOR SEQ ID NO: 7:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1068 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-09-051-341-7

Query Match 66.7%; Score 36; DB 3; Length 1068;
Best Local Similarity 66.7%; Pred. No. 68;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMDYS 11
|:|:|:|:
Db 435 VPPGMDFS 443

RESULT 8
US-09-394-272-8
/ Sequence 8, Application US/09394272
/ Patent No. 6472588
/ GENERAL INFORMATION:
/ APPLICANT: Haigler, Candace H.
/ APPLICANT: Holaday, A. Scott
/ TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
/ TITLE OF INVENTION: EXPRESSION OF SUCROSE PHOSPHATE SYNTHASE
/ FILE REFERENCE: 201304/1000
/ CURRENT APPLICATION NUMBER: US/09/394,272
/ CURRENT FILING DATE: 1999-09-10
/ NUMBER OF SEQ ID NOS: 14
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 8
/ LENGTH: 1068
/ TYPE: PRT
/ ORGANISM: Zea mays
/ US-09-394-272-8

Query Match 66.7%; Score 36; DB 4; Length 1068;
Best Local Similarity 66.7%; Pred. No. 68;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMDYS 11
|:|:|:|:
Db 435 VPPGMDFS 443

RESULT 9
US-09-394-272-4
/ Sequence 4, Application US/09394272
/ Patent No. 6472588
/ GENERAL INFORMATION:
/ APPLICANT: Haigler, Candace H.
/ APPLICANT: Holaday, A. Scott
/ TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
/ TITLE OF INVENTION: EXPRESSION OF SUCROSE PHOSPHATE SYNTHASE
/ FILE REFERENCE: 201304/1000
/ CURRENT APPLICATION NUMBER: US/09/394,272
/ CURRENT FILING DATE: 1999-09-10
/ NUMBER OF SEQ ID NOS: 14
/ SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 4
; LENGTH: 1081
; TYPE: PRT
; ORGANISM: Craterostigma plantagineum
US-09-394-272-4

Query Match 66.7%; Score 36; DB 4; Length 1081;
Best Local Similarity 66.7%; Pred. No. 69;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMDYS 11
|:|:|:|:
Db 445 VIPPGMDFS 453

RESULT 10

US-09-394-272-11
; Sequence 11, Application US/09394272
; Patent No. 6472588
; GENERAL INFORMATION:
; APPLICANT: Haigler, Candace H.
; APPLICANT: Holaday, A. Scott
; TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
; TITLE OF INVENTION: EXPRESSION OF SUCROSE PHOSPHATE SYNTHASE
; FILE REFERENCE: 201304/1000
; CURRENT APPLICATION NUMBER: US/09/394,272
; CURRENT FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 1083
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-394-272-11

Query Match 66.7%; Score 36; DB 4; Length 1083;
Best Local Similarity 66.7%; Pred. No. 69;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMDYS 11
|:|:|:|:
Db 483 VIPPGMDFS 491

RESULT 11

US-09-394-272-9
; Sequence 9, Application US/09394272
; Patent No. 6472588
; GENERAL INFORMATION:
; APPLICANT: Haigler, Candace H.
; APPLICANT: Holaday, A. Scott
; TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
; TITLE OF INVENTION: EXPRESSION OF SUCROSE PHOSPHATE SYNTHASE
; FILE REFERENCE: 201304/1000
; CURRENT APPLICATION NUMBER: US/09/394,272
; CURRENT FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1084
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-394-272-9

Query Match 66.7%; Score 36; DB 4; Length 1084;
Best Local Similarity 66.7%; Pred. No. 69;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMDYS 11
|:|:|:|:
Db 453 VIPPGMDFS 461

RESULT 12

US-08-569-147-76
; Sequence 76, Application US/08569147
; Patent No. 6180377
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: HUMANISED ANTIBODIES
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
; ADDRESSEE: No. 6180377rie, LLP
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/569,147
; FILING DATE: 25-March-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Trujillo, Doreen Yanko
; REGISTRATION NUMBER: 35,719
; REFERENCE/DOCKET NUMBER: CARP-0047
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 76:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 140 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-569-147-76

Query Match 63.0%; Score 34; DB 3; Length 140;
Best Local Similarity 75.0%; Pred. No. 16;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 VVPXGMDY 10
|:|:|:|:
Db 122 WPTGFDY 129

RESULT 13

US-08-569-147-82
; Sequence 82, Application US/08569147
; Patent No. 6180377
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: HUMANISED ANTIBODIES
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
; ADDRESSEE: No. 6180377rie, LLP
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/569,147
; FILING DATE: 25-March-1996

; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Trujillo, Doreen Yanko
; REGISTRATION NUMBER: 35,719
; REFERENCE/DOCKET NUMBER: CARP-0047
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 82:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 140 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-569-147-82

Query Match 63.0%; Score 34; DB 3; Length 140;
Best Local Similarity 75.0%; Pred. No. 16;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGMDY 10
| | | | |
DB 122 VVPTGFDY 129

RESULT 14

US-09-252-991A-31637
; Sequence 31637, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31637
; LENGTH: 1065
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-31637

Query Match 63.0%; Score 34; DB 4; Length 1065;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 PXGMDYS 11
| | | | |
DB 324 PQGMDYS 330

RESULT 15

US-08-963-851-14
; Sequence 14, Application US/08963851
; Patent No. 6300116
; GENERAL INFORMATION:
; APPLICANT: VAN DER OSTEN, CLAUS
; APPLICANT: HALKIER, TORDEN
; APPLICANT: ANDERSEN, CARSTEN
; APPLICANT: BAUDITZ, PETER
; APPLICANT: HANSEN, PETER KAMP
; TITLE OF INVENTION: PROTEASE VARIANTS AND COMPOSITIONS
; FILE REFERENCE: 4946,200-US
; CURRENT APPLICATION NUMBER: US/08/963.851
; CURRENT FILING DATE: 1997-11-04
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 14

; LENGTH: 59
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
; US-08-963-851-14

Query Match 61.1%; Score 33; DB 4; Length 59;
Best Local Similarity 45.5%; Pred. NO. 9.6;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMDYS 11
| : : | : : |
DB 38 EXHIPGGLYS 48

Search completed: December 22, 2003, 16:43:44
Job time : 11.1333 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:40:14 ; Search time 20.8667 Seconds
(without alignments)
98.451 Million cell updates/sec

Title: US-09-909-164-13

Perfect score: 54

Sequence: 1 BEVVPXGMDYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 696363 seqs, 186758610 residues

Total number of hits satisfying chosen parameters: 696363

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 36 | 66.7 | 126 | 12 | US-10-393-840-946 |
| 2 | 36 | 66.7 | 937 | 12 | US-10-289-757-73 |
| 3 | 36 | 66.7 | 938 | 12 | US-10-289-757-168 |
| 4 | 36 | 66.7 | 1049 | 15 | US-10-217-700-10 |
| 5 | 36 | 66.7 | 1062 | 12 | US-10-289-757-71 |
| 6 | 36 | 66.7 | 1068 | 15 | US-10-217-700-8 |
| 7 | 36 | 66.7 | 1081 | 15 | US-10-217-700-4 |
| 8 | 36 | 66.7 | 1083 | 15 | US-10-217-700-11 |
| 9 | 36 | 66.7 | 1084 | 15 | US-10-217-700-9 |
| 10 | 35 | 64.8 | 440 | 11 | US-09-813-408-27 |
| 11 | 34.5 | 63.9 | 588 | 12 | US-10-094-749-1660 |
| 12 | 34 | 63.0 | 222 | 12 | US-10-091-007-58 |
| 13 | 34 | 63.0 | 1062 | 9 | US-09-815-242-5111 |
| 14 | 34 | 63.0 | 3472 | 12 | US-10-029-120-4 |
| 15 | 34 | 63.0 | 3472 | 14 | US-10-027-806-4 |

| | | | | | | |
|----|----|------|------|----|---------------------|-------------------|
| 16 | 34 | 63.0 | 3472 | 14 | US-10-034-623-4 | Sequence 4, Appli |
| 17 | 34 | 63.0 | 3472 | 15 | US-10-027-801-4 | Sequence 4, Appli |
| 18 | 33 | 61.1 | 59 | 10 | US-09-948-080-14 | Sequence 14, Appl |
| 19 | 33 | 61.1 | 236 | 12 | US-10-029-386-32076 | Sequence 32076, A |
| 20 | 33 | 61.1 | 283 | 10 | US-09-738-626-4881 | Sequence 4881, A |
| 21 | 33 | 61.1 | 299 | 9 | US-09-815-242-10697 | Sequence 10697, A |
| 22 | 33 | 61.1 | 322 | 15 | US-10-156-761-12418 | Sequence 12418, A |
| 23 | 33 | 61.1 | 736 | 10 | US-09-978-295A-526 | Sequence 526, App |
| 24 | 33 | 61.1 | 736 | 10 | US-09-978-697-526 | Sequence 526, App |
| 25 | 33 | 61.1 | 736 | 10 | US-09-978-192A-526 | Sequence 526, App |
| 26 | 33 | 61.1 | 736 | 10 | US-09-999-832A-526 | Sequence 526, App |
| 27 | 33 | 61.1 | 736 | 11 | US-09-978-189-526 | Sequence 526, App |
| 28 | 33 | 61.1 | 736 | 11 | US-09-978-608A-526 | Sequence 526, App |
| 29 | 33 | 61.1 | 736 | 11 | US-09-978-585A-526 | Sequence 526, App |
| 30 | 33 | 61.1 | 736 | 11 | US-09-978-191A-526 | Sequence 526, App |
| 31 | 33 | 61.1 | 736 | 11 | US-09-978-403A-526 | Sequence 526, App |
| 32 | 33 | 61.1 | 736 | 11 | US-09-978-564A-526 | Sequence 526, App |
| 33 | 33 | 61.1 | 736 | 11 | US-09-999-833A-526 | Sequence 526, App |
| 34 | 33 | 61.1 | 736 | 11 | US-09-981-915A-526 | Sequence 526, App |
| 35 | 33 | 61.1 | 736 | 11 | US-09-978-824-526 | Sequence 526, App |
| 36 | 33 | 61.1 | 736 | 11 | US-09-918-585A-526 | Sequence 526, App |
| 37 | 33 | 61.1 | 736 | 11 | US-09-978-423A-526 | Sequence 526, App |
| 38 | 33 | 61.1 | 736 | 11 | US-09-978-193A-526 | Sequence 526, App |
| 39 | 33 | 61.1 | 736 | 11 | US-09-999-830A-526 | Sequence 526, App |
| 40 | 33 | 61.1 | 736 | 11 | US-09-978-757A-526 | Sequence 526, App |
| 41 | 33 | 61.1 | 736 | 11 | US-09-978-187B-526 | Sequence 526, App |
| 42 | 33 | 61.1 | 736 | 11 | US-09-978-643A-526 | Sequence 526, App |
| 43 | 33 | 61.1 | 736 | 12 | US-09-978-375A-526 | Sequence 526, App |
| 44 | 33 | 61.1 | 736 | 12 | US-09-978-188A-526 | Sequence 526, App |
| 45 | 33 | 61.1 | 736 | 12 | US-09-978-298A-526 | Sequence 526, App |

ALIGNMENTS

RESULT 1

US-10-393-840-946
; Sequence 946, Application US/10393840
; Publication No. US20030229922A1
; GENERAL INFORMATION:
; APPLICANT: Bloksberg, Leonard N.
; TITLE OF INVENTION: Materials and Methods for the
; Publication of Plant Cell Wall Polysaccharides
; FILE REFERENCE: 11000.1012c3
; CURRENT APPLICATION NUMBER: US/10/393,840
; CURRENT FILING DATE: 2003-03-20
; PRIOR APPLICATION NUMBER: US 09/636,800
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: US 09/170,862
; PRIOR FILING DATE: 1998-10-13
; PRIOR APPLICATION NUMBER: US 60/148,426
; PRIOR FILING DATE: 1999-08-11
; PRIOR APPLICATION NUMBER: PCT NZ/99/00169
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 956
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 946
; LENGTH: 126
; TYPE: PRT
; ORGANISM: Pinus radiata
US-10-393-840-946

Query Match 66.7%; Score 36; DB 12; Length 126;

Best Local Similarity 66.7%; Pred. No. 14; Indels 0; Gaps 0;
Matches 6; Conservative 2; Mismatches 1;

Qv 3 VVFXGMDYS 11

|:|:|:|

Db 23 VVFXGMDYS 31

RESULT 2

US-10-289-757-73

Sequence 73, Application US/10289757
 ; Publication No. US20030180751A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Demmer, Jerroen
 ; APPLICANT: Forster, Richard L
 ; APPLICANT: Gibson, John Bryan
 ; APPLICANT: Shenk, Michael Andrew
 ; APPLICANT: No. US20030180751A1riss, Geoffrey
 ; APPLICANT: Glenn, Matthew
 ; APPLICANT: Saulsbury, Keith Martin
 ; APPLICANT: Hall, Claire
 ; TITLE OF INVENTION: Compositions isolated from forage
 ; TITLE OF INVENTION: grasses and methods for their use
 ; FILE REFERENCE: 11000.1061U
 ; CURRENT APPLICATION NUMBER: US/10/289,757
 ; PRIOR FILING DATE: 2002-11-07
 ; PRIOR FILING DATE: 2001-11-07
 ; NUMBER OF SEQ ID NOS: 218
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 73
 ; LENGTH: 937
 ; TYPE: PRT
 ; ORGANISM: Lolium perenne
 US-10-289-757-73

Query Match 66.7%; Score 36; DB 12; Length 937;
 Best Local Similarity 66.7%; Pred. No. 1.3e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMDYS 11
 |:|:|:|:
 Db 462 VIPPGMDFS 470

RESULT 3

US-10-289-757-168
 ; Sequence 168, Application US/10289757
 ; Publication No. US20030180751A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Demmer, Jerroen
 ; APPLICANT: Forster, Richard L
 ; APPLICANT: Gibson, John Bryan
 ; APPLICANT: Shenk, Michael Andrew
 ; APPLICANT: No. US20030180751A1riss, Geoffrey
 ; APPLICANT: Glenn, Matthew
 ; APPLICANT: Saulsbury, Keith Martin
 ; APPLICANT: Hall, Claire
 ; TITLE OF INVENTION: Compositions isolated from forage
 ; TITLE OF INVENTION: grasses and methods for their use
 ; FILE REFERENCE: 11000.1061U
 ; CURRENT APPLICATION NUMBER: US/10/289,757
 ; PRIOR FILING DATE: 2002-11-07
 ; PRIOR FILING DATE: 2001-11-07
 ; NUMBER OF SEQ ID NOS: 218
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 168
 ; LENGTH: 938
 ; TYPE: PRT
 ; ORGANISM: Lolium perenne
 US-10-289-757-168

Query Match 66.7%; Score 36; DB 12; Length 938;
 Best Local Similarity 66.7%; Pred. No. 1.3e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMDYS 11
 |:|:|:|:
 Db 462 VIPPGMDFS 470

RESULT 4

US-10-217-700-10
 ; Sequence 10, Application US/10217700
 ; Publication No. US20030070191A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haigler, Candace H.
 ; APPLICANT: Holaday, A. Scott
 ; TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
 ; TITLE OF INVENTION: EXPRESSION OF SUCROSE PHOSPHATE SYNTHASE
 ; FILE REFERENCE: 201304/1000
 ; CURRENT APPLICATION NUMBER: US/10/217,700
 ; CURRENT FILING DATE: 2002-08-12
 ; EARLIER APPLICATION NUMBER: 09/394,272
 ; EARLIER FILING DATE: 1999-09-10
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 10
 ; LENGTH: 1049
 ; TYPE: PRT
 ; ORGANISM: Oryza sativa
 US-10-217-700-10

Query Match 66.7%; Score 36; DB 15; Length 1049;
 Best Local Similarity 66.7%; Pred. No. 1.5e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMDYS 11
 |:|:|:|:
 Db 436 VIPPGMDFS 444

RESULT 5

US-10-289-757-71
 ; Sequence 71, Application US/10289757
 ; Publication No. US20030180751A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Demmer, Jerroen
 ; APPLICANT: Forster, Richard L
 ; APPLICANT: Gibson, John Bryan
 ; APPLICANT: Shenk, Michael Andrew
 ; APPLICANT: No. US20030180751A1riss, Geoffrey
 ; APPLICANT: Glenn, Matthew
 ; APPLICANT: Saulsbury, Keith Martin
 ; APPLICANT: Hall, Claire
 ; TITLE OF INVENTION: Compositions isolated from forage
 ; TITLE OF INVENTION: grasses and methods for their use
 ; FILE REFERENCE: 11000.1061U
 ; CURRENT APPLICATION NUMBER: US/10/289,757
 ; PRIOR FILING DATE: 2002-11-07
 ; PRIOR FILING DATE: 2001-11-07
 ; NUMBER OF SEQ ID NOS: 218
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 71
 ; LENGTH: 1062
 ; TYPE: PRT
 ; ORGANISM: Festuca arundinacea
 US-10-289-757-71

Query Match 66.7%; Score 36; DB 12; Length 1062;
 Best Local Similarity 66.7%; Pred. No. 1.5e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMDYS 11
 |:|:|:|:
 Db 466 VIPPGMDFS 474

RESULT 6

US-10-217-700-8
 ; Sequence 8, Application US/10217700
 ; Publication No. US20030070191A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haigler, Candace H.

; APPLICANT: Holaday, A. Scott
; TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
; FILE REFERENCE: 201304/1000
; CURRENT APPLICATION NUMBER: US/10/217,700
; CURRENT FILING DATE: 2002-08-12
; EARLIER APPLICATION NUMBER: 09/394,272
; EARLIER FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1068
; TYPE: PRT
; ORGANISM: Zea mays
US-10-217-700-8

Query Match 66.7%; Score 36; DB 15; Length 1068;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMDYS 11
|:|:|:|:|:
Db 435 VIPPGMDFS 443

RESULT 7
US-10-217-700-4
; Sequence 4, Application US/10217700
; Publication No. US20030070191A1
; GENERAL INFORMATION:
; APPLICANT: Haigler, Candace H.
; APPLICANT: Holaday, A. Scott
; TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
; FILE REFERENCE: 201304/1000
; CURRENT APPLICATION NUMBER: US/10/217,700
; CURRENT FILING DATE: 2002-08-12
; EARLIER APPLICATION NUMBER: 09/394,272
; EARLIER FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1081
; TYPE: PRT
; ORGANISM: Craterostigma plantagineum
US-10-217-700-4

Query Match 66.7%; Score 36; DB 15; Length 1081;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMDYS 11
|:|:|:|:|:
Db 445 VIPPGMDFS 453

RESULT 8
US-10-217-700-11
; Sequence 11, Application US/10217700
; Publication No. US20030070191A1
; GENERAL INFORMATION:
; APPLICANT: Haigler, Candace H.
; APPLICANT: Holaday, A. Scott
; TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
; FILE REFERENCE: 201304/1000
; CURRENT APPLICATION NUMBER: US/10/217,700
; CURRENT FILING DATE: 2002-08-12
; EARLIER APPLICATION NUMBER: 09/394,272
; EARLIER FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11

; LENGTH: 1083
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-217-700-11

Query Match 66.7%; Score 36; DB 15; Length 1083;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMDYS 11
|:|:|:|:|:
Db 483 VIPPGMDFS 491

RESULT 9
US-10-217-700-9
; Sequence 9, Application US/10217700
; Publication No. US20030070191A1
; GENERAL INFORMATION:
; APPLICANT: Haigler, Candace H.
; APPLICANT: Holaday, A. Scott
; TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
; FILE REFERENCE: 201304/1000
; CURRENT APPLICATION NUMBER: US/10/217,700
; CURRENT FILING DATE: 2002-08-12
; EARLIER APPLICATION NUMBER: 09/394,272
; EARLIER FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1084
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-217-700-9

Query Match 66.7%; Score 36; DB 15; Length 1084;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMDYS 11
|:|:|:|:|:
Db 453 VIPPGMDFS 461

RESULT 10
US-09-813-408-27
; Sequence 27, Application US/09813408
; Publication No. US20030049619A1
; GENERAL INFORMATION:
; APPLICANT: Delagrave, Simon
; APPLICANT: Harris, Barry
; TITLE OF INVENTION: Methods For The Synthesis Of Polynucleotides And Combinatorial
; FILE REFERENCE: HER0041
; CURRENT APPLICATION NUMBER: US/09/813,408
; CURRENT FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 27
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Aeropyrum pernix
US-09-813-408-27

Query Match 64.8%; Score 35; DB 11; Length 440;
Best Local Similarity 66.7%; Pred. No. 92;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 EVVPXGMDY 10
|:|:|:|:|:
Db 120 EVLPWGVGY 128

```

RESULT 11
US-10-094-749-1660
; Sequence 1660, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOTYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1660
; LENGTH: 588
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-1660

Query Match      63.9%; Score 34.5; DB 12; Length 588;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches      8; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

Qy      1 BEVVPXGMDY 10
Db      127 BEVVP-GMDF 135

RESULT 12
US-10-091-007-58
; Sequence 58, Application US/10091007
; Publication No. US20030170782A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics limited
; APPLICANT: Le Page, Richard W F Wells, Jeremy M
; APPLICANT: Hanniffy, Sean B
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21978WO
; CURRENT APPLICATION NUMBER: US/10/091,007
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: GB 9921125.2
; PRIOR FILING DATE: 1999-09-07
; NUMBER OF SEQ ID NOS: 276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 58
; LENGTH: 222
; TYPE: PRT
; ORGANISM: Streptococcus agalactiae
US-10-091-007-58

Query Match      63.0%; Score 34; DB 12; Length 222;
Best Local Similarity 50.0%; Pred. No. 68;

Qy      1 BEVVPXGMDY 10
Db      127 BEVVP-GMDF 135

RESULT 13
US-09-815-242-5111
; Sequence 5111, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5111
; LENGTH: 1062
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-5111

Query Match      63.0%; Score 34; DB 9; Length 1062;
Best Local Similarity 85.7%; Pred. No. 3.9e+02;
Matches      6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      5 PGMMDYS 11
Db      321 PGMMDYS 327

RESULT 14
US-10-029-120-4
; Sequence 4, Application US/10029120
; Publication No. US20030175708A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAUM SYMBIOSUM
; FILE REFERENCE: DCRP.002A
; CURRENT APPLICATION NUMBER: US/10/029,120
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4

```

```
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-029-120-4

Query Match      63.0%; Score 34; DB 12; Length 3472;
Best Local Similarity 45.5%; Pred. No. 1.5e+03;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY      1 EEVVPXGMDYS 11
      |:|:|:|:|:|
Db      2294 EDVIPRGISFS 2304

RESULT 15
US-10-027-806-4
; Sequence 4, Application US/10027806
; Publication No. US20020160476A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCOF.002A
; CURRENT APPLICATION NUMBER: US/10/027,806
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 3472
; TYPE: PRT
; ORGANISM: Cenarchaeum symbiosum
US-10-027-806-4

Query Match      63.0%; Score 34; DB 14; Length 3472;
Best Local Similarity 45.5%; Pred. No. 1.5e+03;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY      1 EEVVPXGMDYS 11
      |:|:|:|:|:|
Db      2294 EDVIPRGISFS 2304
```

Search completed: December 22, 2003, 17:32:43
Job time : 20.9333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 17:24:36 ; Search time 9.06667 Seconds
(without alignments)
116.675 Million cell updates/sec

Title: US-09-909-164-13

Perfect score: 54

Sequence: 1 EVVFXGMDYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Pir1.*
2: Pir2.*
3: Pir3.*
4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|----------|--------------------|
| 1 | 40 | 74.1 | 156 | 2 S54619 | hypothetical prote |
| 2 | 38 | 70.4 | 363 | 2 D69551 | conserved hypotet |
| 3 | 36 | 66.7 | 102 | 2 A42452 | V1 protein - tobac |
| 4 | 36 | 66.7 | 341 | 2 S72649 | sucrose-phosphate |
| 5 | 36 | 66.7 | 348 | 2 S72650 | sucrose-phosphate |
| 6 | 36 | 66.7 | 460 | 2 G96784 | unknown protein F2 |
| 7 | 36 | 66.7 | 1049 | 2 JC4783 | sucrose-phosphate |
| 8 | 36 | 66.7 | 1068 | 1 JQ1329 | sucrose-phosphate |
| 9 | 36 | 66.7 | 1081 | 2 T09837 | sucrose-phosphate |
| 10 | 36 | 66.7 | 1083 | 2 T04062 | sucrose-phosphate |
| 11 | 36 | 66.7 | 1084 | 2 T04103 | sucrose-phosphate |
| 12 | 35 | 64.8 | 425 | 2 T24111 | hypothetical prote |
| 13 | 35 | 64.8 | 433 | 2 H87660 | peptidoglycan-bind |
| 14 | 35 | 64.8 | 440 | 2 H72784 | probable alkaline |
| 15 | 35 | 64.8 | 1150 | 2 T20173 | hypothetical prote |
| 16 | 35 | 64.8 | 1474 | 2 P69009 | probable membrane |
| 17 | 35 | 64.8 | 2747 | 2 B49132 | fat facets (faf) a |
| 18 | 34 | 63.0 | 99 | 2 S00210 | plastocyanin b - L |
| 19 | 34 | 63.0 | 155 | 2 S38255 | plastocyanin precu |
| 20 | 34 | 63.0 | 168 | 2 S58208 | plastocyanin b pre |
| 21 | 34 | 63.0 | 290 | 2 AG3104 | 6-O-methylguanine- |
| 22 | 34 | 63.0 | 290 | 2 D98182 | O6-methylguanine-D |
| 23 | 34 | 63.0 | 296 | 2 F72745 | hypothetical prote |
| 24 | 34 | 63.0 | 357 | 1 G69290 | probable hexosyltr |
| 25 | 34 | 63.0 | 366 | 2 G69350 | L-lactate dehydrog |
| 26 | 34 | 63.0 | 565 | 2 E86665 | ABC transporter AT |
| 27 | 34 | 63.0 | 566 | 2 A70164 | phenylalanine-tRNA |
| 28 | 34 | 63.0 | 587 | 2 P81138 | succinate dehydrog |
| 29 | 34 | 63.0 | 906 | 2 T48898 | disease resistance |

| | | | | | |
|----|----|------|------|----------|---------------------|
| 30 | 34 | 63.0 | 908 | 2 T48899 | disease resistance |
| 31 | 34 | 63.0 | 1062 | 2 P83335 | RND multidrug effl |
| 32 | 34 | 63.0 | 1082 | 2 T30830 | hypothetical prote |
| 33 | 34 | 63.0 | 3472 | 2 T31308 | hypothetical 367K |
| 34 | 33 | 61.1 | 97 | 2 A9427 | partial transposase |
| 35 | 33 | 61.1 | 128 | 2 A90471 | hypothetical prote |
| 36 | 33 | 61.1 | 172 | 2 S27021 | fibroblast growth |
| 37 | 33 | 61.1 | 184 | 2 S90335 | hypothetical prote |
| 38 | 33 | 61.1 | 225 | 2 S57810 | hypothetical prote |
| 39 | 33 | 61.1 | 247 | 2 A96001 | conserved hypotet |
| 40 | 33 | 61.1 | 257 | 2 A96546 | unknown protein [i |
| 41 | 33 | 61.1 | 262 | 2 F90298 | transposase ISC105 |
| 42 | 33 | 61.1 | 267 | 2 C90307 | transposase ISC105 |
| 43 | 33 | 61.1 | 276 | 2 C64417 | hypothetical prote |
| 44 | 33 | 61.1 | 283 | 2 G83055 | pantoate-beta-alan |
| 45 | 33 | 61.1 | 299 | 2 E90487 | transposase ISC105 |

ALIGNMENTS

RESULT 1

S54619
hypothetical protein YOR013w - yeast (Saccharomycetes cerevisiae)
N;Alternate names: hypothetical protein O2612; hypothetical protein YOL303.3
C;Species: Saccharomycetes cerevisiae
C;Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002
C;Accession: S54619; S66879
R;de Haan, M.; Maarsee, A.C.; Grivell, L.A.
submitted to the EMBL Data Library, May 1995
A;Reference number: S54617
A;Accession: S54619
A;Molecule type: DNA
A;Residues: 1-156 <DEH>
A;Cross-references: EMBL:X87331; NID:g1041652; PIDN:CAA60762.1; PID:g829123
R;de Haan, M.; Grivell, L.A.; Maarsee, A.C.
submitted to the Protein Sequence Database, July 1996
A;Reference number: S66877
A;Accession: S66879
A;Molecule type: DNA
A;Residues: 1-156 <DEW>
A;Cross-references: EMBL:Z74920; NID:g1420109; PIDN:CAA99201.1; PID:g1420111; MIPS:YOR
A;Experimental source: strain S288C
C;Genetics:
A;Cross-references: SGD:S0005539
A;Map position: 15R
C;Superfamily: hypothetical protein YOR013w

Query Match 74.1%; Score 40; DB 2; Length 156;
Best Local Similarity 77.8%; Pred. No. 1;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 EVVFXGMDY 10
||:||||
Db 50 EVMPLGMDY 58

RESULT 2

D69551
conserved hypothetical protein AF2411 - Archaeoglobus fulgidus
C;Species: Archaeoglobus fulgidus
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C;Accession: D69551
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.P.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dods;
; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.P.; McDonald, L.
Nature 390, 364-370, 1997
A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archa
A;Reference number: A69250; MUID:98049343; PMID:9389475
A;Accession: D69551
A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA
A:Residues: 1-363 <KLE>
A:Cross-references: GB:AE001109; GB:AE000782; NID:G2609432; PIDN:AAB91255.1; PID:G265068

Query Match 70.4%; Score 38; DB 2; Length 363;
Best Local Similarity 54.5%; Pred. No. 6.8;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 BEVVPXGMDYS 11
| : || : || : ||
Db 120 ENIVPYGIDFS 130

RESULT 3
A42452
V1 protein - tobacco yellow dwarf virus (strain Australia)
C:Species: tobacco yellow dwarf virus
C:Date: 15-Jan-1993 #sequence_revision 15-Jan-1993 #text_change 08-Oct-1999
C:Accession: A42452
R:Morris, B.A.M.; Richardson, K.A.; Haley, A.; Zhan, X.; Thomas, J.E.
Virology 187, 633-642, 1992
A:Title: The nucleotide sequence of the infectious cloned DNA component of tobacco yellow
A:Reference number: A42452; MUID:92188538; PMID:1546458
A:Accession: A42452
A:Molecule type: DNA
A:Residues: 1-102 <MOR>
A:Cross-references: GB:M81103; NID:G335283; PIDN:AAA47947.1; PID:G335284

Query Match 66.7%; Score 36; DB 2; Length 102;
Best Local Similarity 60.0%; Pred. No. 4.3;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 2 EVVPXGMDYS 11
: ||| : ||| : |||
Db 7 QVVPFGINYS 16

RESULT 4
S72649
sucrose-phosphate synthase (EC 2.4.1.14) isoform 2 - Citrus unshiu (fragment)
C:Species: Citrus unshiu
C:Date: 24-Oct-1998 #sequence_revision 24-Oct-1998 #text_change 21-Jul-2000
C:Accession: S72649
R:Komatsu, A.; Takanokura, Y.; Omura, M.; Akihama, T.
Mol. Gen. Genet. 252, 346-351, 1996
A:Title: Cloning and molecular analysis of cDNAs encoding three sucrose phosphate synthase
A:Reference number: S72648; MUID:96439842; PMID:8842155
A:Accession: S72649
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-341 <KOM>
A:Cross-references: EMBL:AB006319; NID:G2588891; PIDN:BAA23215.1; PID:G2588892
A:Experimental source: fruit, cv. Miyagawa-Wase
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1997
C:Genetics:
A:Gene: SPS2
C:Function:
A:Description: catalyzes formation of sucrose-6-phosphate from UDPglucose and D-fructose
A:Pathway: sucrose biosynthesis
C:Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology
C:Keywords: glycosyltransferase; hexosyltransferase; sucrose biosynthesis
F:1-341/Domain: sucrose/sucrose-phosphate synthase homology (fragment) <SPS>

Query Match 66.7%; Score 36; DB 2; Length 341;
Best Local Similarity 66.7%; Pred. No. 16;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMDYS 11
| : || : || : ||
Db 228 VIPPGMDFS 236

RESULT 5
S72650
sucrose-phosphate synthase (EC 2.4.1.14) isoform 3 - Citrus unshiu (fragment)
C:Species: Citrus unshiu
C:Date: 24-Oct-1998 #sequence_revision 24-Oct-1998 #text_change 21-Jul-2000
C:Accession: S72650
R:Komatsu, A.; Takanokura, Y.; Omura, M.; Akihama, T.
Mol. Gen. Genet. 252, 346-351, 1996
A:Title: Cloning and molecular analysis of cDNAs encoding three sucrose phosphate synthase
A:Reference number: S72648; MUID:96439842; PMID:8842155
A:Accession: S72650
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-348 <KOM>
A:Cross-references: EMBL:AB006660; NID:G2351059; PIDN:BAA22071.1; PID:G2351060
A:Experimental source: fruit, cv. Miyagawa-Wase
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1997
C:Genetics:
A:Gene: SPS3
C:Function:
A:Description: catalyzes formation of sucrose-6-phosphate from UDPglucose and D-fructose
A:Pathway: sucrose biosynthesis
C:Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology
C:Keywords: glycosyltransferase; hexosyltransferase; sucrose biosynthesis
F:1-348/Domain: sucrose/sucrose-phosphate synthase homology (fragment) <SPS>

Query Match 66.7%; Score 36; DB 2; Length 348;
Best Local Similarity 66.7%; Pred. No. 17;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMDYS 11
| : || : || : ||
Db 234 VIPPGMDFS 242

RESULT 6
G96764
unknown protein F25P22.17 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: G96764
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alons
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K
ansen, N.P.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzial
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: G96764
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-460 <SPO>
A:Cross-references: GB:AE005173; NID:G6692750; PIDN:AAF24856.1; GSPDB:GN00141
C:Genetics:
A:Gene: F25P22.17
A:Map position: 1

Query Match 66.7%; Score 36; DB 2; Length 460;
Best Local Similarity 70.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVVPXGMDY 10
| : || : || : ||
Db 218 BEVVPXGMDY 227

RESULT 7
JC4783
sucrose-phosphate synthase (EC 2.4.1.14) - rice
C:Species: Oryza sativa (rice)

C>Date: 10-May-1996 #sequence_revision 16-Aug-1996 #text_change 18-Jun-1999
 C/Accession: J04783
 R/Valdez-Alarcon, J.J.; Ferrando, M.; Salerno, G.; Jimenez-Moraila, B.; Herrera-Bastrelle
 Gene 170, 217-222, 1996
 A/Title: Characterization of a rice sucrose-phosphate synthase-encoding gene.
 A/Reference number: J04783; MUID:96235138; PMID:8666248
 A/Accession: J04783
 A/Molecule type: mRNA
 A/Residues: 1-1049 <VAL>
 A/Cross-references: GB:U33175; NID:g1449931; PIDN:AAC49379.1; PID:g988270
 A/Note: UDPglucose-fructose-phosphate glucosyltransferase; Sucrosephosphate-UDPglucosyl
 C/Comment: This enzyme catalyzes the formation of sucrose-phosphate from UDP-glucose and
 C/Genetics: JC4783
 A/Gene: Spel
 A/Introns: 24/1, 103/3; 183/3; 205/3; 435/3; 475/3; 519/3; 578/3; 596/3; 617/3; 931/3; 9
 C/Function:
 A/Description: catalyzes the formation of sucrose-6-phosphate from UDPglucose and D-fruc
 A/Pathway: sucrose biosynthesis
 C/Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology
 C/Keywords: glycosyltransferase; hexosyltransferase; sucrose biosynthesis
 F:180-663/Domain: sucrose/sucrose-phosphate synthase homology <SSPS>

Query Match 66.7%; Score 36; DB 2; Length 1049;
 Best Local Similarity 66.7%; Pred. NO. 58;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMDYS 11
 Db 436 VIPPGMDFS 444

RESULT 8
 J01329
 sucrose-phosphate synthase (EC 2.4.1.14) - maize
 C/Species: Zea mays (maize)
 C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C/Accession: J01329; PQ0260
 R/Worrell, A.C.; Bruneau, J.M.; Summerfelt, K.; Boersig, M.; Voelker, T.A.
 Plant Cell 3, 1121-1130, 1991
 A/Title: Expression of a maize sucrose phosphate synthase in tomato alters leaf carbohyd
 A/Reference number: J01329; MUID:92338837; PMID:1840396
 A/Accession: J01329
 A/Molecule type: mRNA
 A/Residues: 1-1068 <WOR>
 A/Cross-references: GB:M97550; NID:g168625; PIDN:AAA33513.1; PID:g168626
 A/Accession: PQ0260
 A/Molecule type: protein
 A/Residues: 71-74;206-212;471-481;872-892 <WOR1>
 C/Comment: This enzyme transfers the glucosyl group from UDPglucose to fructose-6-phosph
 C/Comment: This enzyme is involved in the regulation of carbon partitioning in the leave
 C/Function:
 A/Description: catalyzes the formation of sucrose-6-phosphate from UDPglucose and D-fruc
 A/Pathway: sucrose biosynthesis
 C/Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology
 C/Keywords: glycosyltransferase; hexosyltransferase; sucrose biosynthesis
 F:178-666/Domain: sucrose/sucrose-phosphate synthase homology <SSPS>

Query Match 66.7%; Score 36; DB 1; Length 1068;
 Best Local Similarity 66.7%; Pred. NO. 59;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMDYS 11
 Db 435 VIPPGMDFS 443

RESULT 9
 T09837
 sucrose-phosphate synthase (EC 2.4.1.14) isoform 2 - Craterostigma plantagineum
 C/Species: Craterostigma plantagineum
 C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
 C/Accession: T09837
 R/Ingram, J.; Chandler, J.W.; Gallagher, L.; Salamini, F.; Bartels, D.

Plant Physiol. 115, 113-121, 1997
 A/Title: Analysis of cDNA clones encoding sucrose-phosphate synthase in relation to su
 A/Reference number: Z16974; MUID:97451773; PMID:9306694
 A/Accession: T09837
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-1081 <ING>
 A/Cross-references: EMBL:Y11795; NID:g2190349; PIDN:CAA72491.1; PID:g2190350
 A/Experimental source: ABA-treated callus
 C/Genetics:
 A/Gene: sps2
 C/Function:
 A/Description: catalyzes the formation of sucrose-6-phosphate from UDPglucose and D-fr
 A/Pathway: sucrose biosynthesis
 C/Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology
 C/Keywords: glycosyltransferase; hexosyltransferase; sucrose biosynthesis
 F:176-674/Domain: sucrose/sucrose-phosphate synthase homology <SSS>

Query Match 66.7%; Score 36; DB 2; Length 1081;
 Best Local Similarity 66.7%; Pred. NO. 60;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMDYS 11
 Db 445 VIPPGMDFS 453

RESULT 10
 T04062
 sucrose-phosphate synthase homolog F28M11.40 - Arabidopsis thaliana
 C/Species: Arabidopsis thaliana (mouse-ear cress)
 C/Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 16-Jul-1999
 C/Accession: T04062
 R/Byan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F
 submitted to the Protein Sequence Database, March 1999
 A/Reference number: Z15184
 A/Accession: T04062
 A/Molecule type: DNA
 A/Residues: 1-1083 <BEV>
 A/Cross-references: EMBL:AL049487
 A/Experimental source: cultivar Columbia; BAC clone F28M11
 C/Genetics:
 A/Map position: 4
 A/Introns: 86/3; 116/3; 255/3; 322/2; 482/3; 526/3; 570/3; 629/3; 647/3; 668/3; 949/3;
 A/Note: F28M11.40
 C/Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology
 F:230-714/Domain: sucrose/sucrose-phosphate synthase homology <SSPS>

Query Match 66.7%; Score 36; DB 2; Length 1083;
 Best Local Similarity 66.7%; Pred. NO. 60;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMDYS 11
 Db 483 VIPPGMDFS 491

RESULT 11
 T04103
 sucrose-phosphate synthase (EC 2.4.1.14) 1 - rice
 C/Species: Oryza sativa (rice)
 C/Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Jun-2000
 C/Accession: T04103
 R/Sakamoto, M.; Satorawa, T.; Kishimoto, N.; Higo, K.; Shimada, H.; Fujimura, T.
 Plant Sci. 112, 207-217, 1995
 A/Title: Structure and RFLP mapping of a rice sucrose phosphate synthase (SPS) gene th
 A/Reference number: Z15212
 A/Accession: T04103
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-1084 <SAK>
 A/Cross-references: EMBL:D45890; PIDN:BAA08304.1
 A/Experimental source: subsp. Japonica

C;Genetics:
A;Gene: Spe1

A;Map position: 1
A;Introns: 120/3; 200/2; 221/3; 452/3; 492/3; 536/3; 595/3; 613/3; 634/3; 946/3; 989/2
C;Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology
C;Keywords: glycosyltransferase; hexosyltransferase
F;156-680/domain: sucrose/sucrose-phosphate synthase homology <SSPS>

Query Match 66.7%; Score 36; DB 2; Length 1084;
Best Local Similarity 66.7%; Pred. No. 60;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMDYS 11
|:|:|:|:|:|:
Db 453 VVPGGMDFS 461

RESULT 12

T24111

hypothetical protein R10D12.10 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T24111

R;Percy, C.

submitted to the EMBL Data Library, October 1996

A;Reference number: Z19842

A;Accession: T24111

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-425 <WIL>

A;Cross-references: EMBL:Z81109; PIDN:CAB03241.1; GSPDB:GN00023; CESP:R10D12.10

A;Experimental source: clone R10D12

C;Genetics:

A;Gene: CESP:R10D12.10

A;Map position: 5

A;Introns: 23/3; 56/3; 113/3; 257/2

Query Match 64.8%; Score 35; DB 2; Length 425;
Best Local Similarity 50.0%; Pred. No. 34;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 BEVVPXGMDY 10
|:|:|:|:|:
Db 335 EQIVPGGLQY 344

RESULT 13

H87660

peptidoglycan-binding protein, probable [imported] - Caulobacter crescentus

C;Species: Caulobacter crescentus

C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001

C;Accession: H87660

R;Niernman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.

B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon

n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A;Title: Complete Genome Sequence of Caulobacter crescentus.

A;Reference number: A87249; MUID:21173698; PMID:11259647

A;Accession: H87660

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-433 <STO>

A;Cross-references: GB:AS005673; NID:gl3425020; PIDN:AAK25284.1; GSPDB:GN00148

C;Genetics:

A;Gene: CC3322

Query Match 64.8%; Score 35; DB 2; Length 433;
Best Local Similarity 54.5%; Pred. No. 34;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVVPXGMDYS 11
|:|:|:|:|:|:
Db 266 EVILPFGFDYS 276

RESULT 14

H72784

probable alkaline proteinase APE0263 - Aeropyrum pernix (strain K1)

C;Species: Aeropyrum pernix

C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000

C;Accession: H72784

R;Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Tak

awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.;

DNA Res. 6, 93-101, 1999

A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aerop

A;Reference number: A72450; MUID:99310339; PMID:10382966

A;Accession: H72784

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-440 <KAW>

A;Cross-references: DDBJ:AP000058; NID:g5103388; PIDN:BAA79178.1; PID:g5103657

A;Experimental source: strain K1

C;Genetics:

A;Gene: APE0263

C;Superfamily: subtilisin; subtilisin homology

Query Match 64.8%; Score 35; DB 2; Length 440;
Best Local Similarity 66.7%; Pred. No. 35;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 EVVPXGMDY 10
|:|:|:|:|:
Db 120 EVLPWGVGY 128

RESULT 15

T20173

hypothetical protein CS3A5.2 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999

C;Accession: T20173; T23857

R;Mortimore, B.

submitted to the EMBL Data Library, November 1996

A;Reference number: Z19232

A;Accession: T20173

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-1150 <WIL>

A;Cross-references: EMBL:Z81486; PIDN:CAB03994.1; GSPDB:GN00023; CESP:CS3A5.2

A;Experimental source: clone CS3A5

R;Matthews, L.

submitted to the EMBL Data Library, August 1996

A;Reference number: Z19808

A;Accession: T23857

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-1150 <WIL>

A;Cross-references: EMBL:Z78015; PIDN:CAB01437.1; GSPDB:GN00023; CESP:CS3A5.2

A;Experimental source: clone R02D5

C;Genetics:

A;Gene: CESP:CS3A5.2

A;Map position: 5

A;Introns: 33/3; 63/3; 132/3; 169/3; 221/3; 299/3; 379/2; 423/2; 438/2; 471/1; 513/2;

Query Match 64.8%; Score 35; DB 2; Length 1150;
Best Local Similarity 66.7%; Pred. No. 1e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPXGMDYS 11
|:|:|:|:|:|:
Db 562 VLPVGDYS 570

Search completed: December 22, 2003, 17:44:58
Job time : 10.0667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:43:51; Search time 4.6 Seconds

(without alignments)
112.455 Million cell updates/sec

Title: US-09-909-164-13

Perfect score: 54

Sequence: 1 EHVVPXGMDYS 11

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|---------------------|
| 1 | 38 | 70.4 | 1058 | 1 | Q8RG86 fusobacteri |
| 2 | 36 | 66.7 | 102 | 1 | P31619 tobacco yel |
| 3 | 36 | 66.7 | 1049 | 1 | Q43802 oryza sativ |
| 4 | 36 | 66.7 | 1068 | 1 | P31927 zea mays (m |
| 5 | 36 | 66.7 | 1081 | 1 | O04933 cratereostig |
| 6 | 35 | 64.8 | 2778 | 1 | P55824 drosophila |
| 7 | 34.5 | 63.9 | 748 | 1 | Q9NR64 homo sapien |
| 8 | 34 | 63.0 | 154 | 1 | P20423 oryza sativ |
| 9 | 34 | 63.0 | 155 | 1 | P08248 hordeum vul |
| 10 | 34 | 63.0 | 168 | 1 | P11970 populus nig |
| 11 | 34 | 63.0 | 566 | 1 | P94283 borrelia bu |
| 12 | 34 | 63.0 | 908 | 1 | Q9FJK8 arabidopsis |
| 13 | 34 | 63.0 | 908 | 1 | Q8W4J9 arabidopsis |
| 14 | 34 | 63.0 | 910 | 1 | P59584 arabidopsis |
| 15 | 33 | 61.1 | 276 | 1 | Q58349 methanococc |
| 16 | 33 | 61.1 | 283 | 1 | Q9HV69 pseudomonas |
| 17 | 33 | 61.1 | 394 | 1 | Q9KMY3 vibrio chol |
| 18 | 33 | 61.1 | 421 | 1 | P08503 rattus norv |
| 19 | 33 | 61.1 | 421 | 1 | O52250 halomonas e |
| 20 | 33 | 61.1 | 423 | 1 | Q3ZEUF halomonas e |
| 21 | 33 | 61.1 | 787 | 1 | O60344 homo sapien |
| 22 | 33 | 61.1 | 801 | 1 | Q61851 mus musculu |
| 23 | 33 | 61.1 | 806 | 1 | P18460 gallus gall |
| 24 | 33 | 61.1 | 877 | 1 | O74377 schizosacch |
| 25 | 33 | 61.1 | 982 | 1 | P27399 simian foam |
| 26 | 33 | 61.1 | 1401 | 1 | Q9KV29 vibrio chol |
| 27 | 33 | 61.1 | 1717 | 1 | P15822 homo sapien |
| 28 | 32.5 | 60.2 | 472 | 1 | P19102 xenopus lae |
| 29 | 32 | 59.3 | 97 | 1 | P20422 daucus caro |
| 30 | 32 | 59.3 | 165 | 1 | Q92BC5 listeria in |
| 31 | 32 | 59.3 | 165 | 1 | Q9Y6U8 listeria mo |
| 32 | 32 | 59.3 | 175 | 1 | Q04667 rattus norv |
| 33 | 32 | 59.3 | 231 | 1 | P08203 escherichia |

RESULT 1

| ID | CARB_FUSNN | STANDARD; | PRT; | 1058 AA. |
|----|--|-----------|------|----------|
| AC | Q8RG86; | | | |
| DT | 28-FEB-2003 (Rel. 41, Created) | | | |
| DT | 28-FEB-2003 (Rel. 41, Last sequence update) | | | |
| DT | 28-FEB-2003 (Rel. 41, Last annotation update) | | | |
| DE | Carbamoyl-phosphate synthase large chain (SC 6.3.5.5) (Carbamoyl-phosphate synthetase ammonia chain). | | | |
| GN | CARB OR FN0422. | | | |
| OS | Fusobacterium nucleatum (subsp. nucleatum). | | | |
| OC | Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae; | | | |
| OC | Fusobacterium. | | | |
| OK | NCBI_TaxID=76856; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN=ATCC 25586; | | | |
| RX | MEDLINE=21886394; PubMed=11889109; | | | |
| RA | Kapatral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A., Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L., Vasieva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A., Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R., Fonstein M., Kyrides N., Overbeek R.; | | | |
| RA | *Genome sequence and analysis of the oral bacterium Fusobacterium nucleatum strain ATCC 25586.*; | | | |
| RL | J. Bacteriol. 184:2005-2018(2002). | | | |
| CC | -1- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP + phosphate + L-glutamate + carbamoyl phosphate. | | | |
| CC | -1- COPACTOR: Binds 3 manganese ions per subunit (By similarity). | | | |
| CC | -1- PATHWAY: Arginine biosynthesis. | | | |
| CC | -1- SUBUNIT: Pyrimidine biosynthesis; first step. | | | |
| CC | -1- PROMOTES: Composed of two chains; the small (or glutamine) chain promotes the hydrolysis of glutamine to ammonia, which is used by the large (or ammonia) chain to synthesize carbamoyl phosphate (By similarity). | | | |
| CC | -1- SIMILARITY: BELONGS TO THE CARB FAMILY. | | | |
| CC | THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch). | | | |
| CC | EMBL; AB010554; AAL94625.1; ALT_INIT. | | | |
| DR | HAWAP; MF_01210; -; 1. | | | |
| DR | InterPro; IPR006275; CarA_L_glu. | | | |
| DR | InterPro; IPR005483; Cphase_L. | | | |
| DR | InterPro; IPR005479; Cphase_L_D2. | | | |
| DR | InterPro; IPR005480; Cphase_L_D3. | | | |
| DR | InterPro; IPR005481; Cphase_L_N. | | | |
| DR | InterPro; IPR004362; MGS-like. | | | |
| DR | Pfam; PF00289; CPSase_L_chain; 2. | | | |
| DR | Pfam; PF02786; CPSase_L_D2; 2. | | | |
| DR | Pfam; PF02787; CPSase_L_D3; 1. | | | |

| | | | | | | |
|----|----|------|-----|---|-------------|--------------------|
| 34 | 32 | 59.3 | 231 | 1 | ARAD_SALTY | P06190 salmonella |
| 35 | 32 | 59.3 | 233 | 1 | HIS9_THEMEA | Q9WZ11 thermotoga |
| 36 | 32 | 59.3 | 288 | 1 | CGD2_RAT | Q04827 rattus norv |
| 37 | 32 | 59.3 | 289 | 1 | CGD2_HUMAN | P30279 homo sapien |
| 38 | 32 | 59.3 | 289 | 1 | CGD2_MOUSE | P30280 mus musculu |
| 39 | 32 | 59.3 | 291 | 1 | CGD1_BRARE | Q90459 brachydanio |
| 40 | 32 | 59.3 | 291 | 1 | CGD1_XENLA | P50755 xenopus lae |
| 41 | 32 | 59.3 | 291 | 1 | CGD2_CHICK | P43702 gallus gall |
| 42 | 32 | 59.3 | 291 | 1 | CGD2_XENLA | P53762 xenopus lae |
| 43 | 32 | 59.3 | 292 | 1 | CGD1_CHICK | P5169 gallus gall |
| 44 | 32 | 59.3 | 292 | 1 | CGD3_HUMAN | P30281 homo sapien |
| 45 | 32 | 59.3 | 295 | 1 | CGD1_HUMAN | P24385 homo sapien |

ALIGNMENTS

```

DR Pfam; PF02142; MGS: 1.
DR PRINTS; PR00098; CFSASE.
DR TIGRFAMs; TIGR01369; CFSaseII_lrg; 1.
DR PROSITE; PS00866; CFSASE_1; 2.
DR PROSITE; PS00867; CFSASE_2; 2.
KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
KW ATP-binding; Manganese; Complete proteome.
FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 402 546 OLIGOMERIZATION DOMAIN.
FT DOMAIN 547 929 CARAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 930 1058 ALLOSTERIC DOMAIN.
FT REPEAT 1 546
FT REPEAT 547 1058
FT NP_BIND 153 210 ATP (POTENTIAL).
FT NP_BIND 302 352 ATP (POTENTIAL).
FT METAL 284 284 MANGANESE 1 (BY SIMILARITY).
FT METAL 298 298 MANGANESE 2 (BY SIMILARITY).
FT METAL 300 300 MANGANESE 3 (BY SIMILARITY).
FT METAL 820 820 MANGANESE 3 (BY SIMILARITY).
FT METAL 832 832 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1058 AA; 117451 MW; ED7037AF77C1E39F CRC64;

Query Match 70.4%; Score 38; DB 1; Length 1058;
Best Local Similarity 60.0%; Pred. No. 9.1;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 2 EVVPGXGMDYS 11
   ||| ||:|
Db 190 EIVPGLNYS 199

RESULT 2
Y1LK TYDVA
ID Y1LK TYDVA STANDARD; PRT; 102 AA.
AC P31619;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DE 01-OCT-1993 (Rel. 27, Last annotation update)
DE Hypothetical 11.2 kDa protein.
GN V1.
OS Tobacco yellow dwarf virus (strain Australia) (TYDV).
OC Viruses; ssDNA viruses; Geminiviridae; Mastrevirus.
OX NCBI_TaxID=31599;
RN SEQUENCE FROM N.A.
RX MEDLINE=92188538; PubMed=1546458;
RA Morris B.A.M., Richardson K.A., Haley A., Zhan X., Thomas J.E.;
RT "The nucleotide sequence of the infectious cloned DNA component of
RT tobacco yellow dwarf virus reveals features of geminiviruses
RT infecting monocotyledonous plants.";
RL Virology 187:633-642(1992).
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL; M81103; AAA47947.1; -.
CC PIR; A42452; A42452.
CC InterPro; IPR002621; Gemini_mov.
CC Pfam; PF01708; Gemini_mov; I.
KW Hypothetical protein.
SQ SEQUENCE 102 AA; 11178 MW; A40ECF1E0AF55B67 CRC64;

Query Match 66.7%; Score 36; DB 1; Length 102;
Best Local Similarity 60.0%; Pred. No. 2;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 2 EVVPGXGMDYS 11
   ||| ||:|

```

```

Db 7 QVVPNGINYS 16

RESULT 3
SPS ORYSA
ID SPS ORYSA STANDARD; PRT; 1049 AA.
AC Q43802;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Sucrose-phosphate synthase (EC 2.4.1.14)
DE (UDP-glucose-fructose-phosphate glucosyltransferase).
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Indica-IR36; TISSUE=leaf;
MEDLINE=96235138; PubMed=866248;
RA Valdez-Alarcon J.J., Ferrando M., Jimenez-Moralla B.,
RA Herrera-Estrella L.;
RT "Characterization of a rice sucrose-phosphate synthase-encoding
RT gene";
RL Gene 170:217-222(1996).
CC -!- FUNCTION: INVOLVED IN THE REGULATION OF CARBON PARTITIONING IN
CC THE LEAVES OF PLANTS. MAY REGULATE THE SYNTHESIS OF SUCROSE AND
CC THEREFORE PLAY A MAJOR ROLE AS A LIMITING FACTOR IN THE EXPORT OF
CC PHOTASSIMILATES OUT OF THE LEAF.
CC -!- CATALYTIC ACTIVITY: UDP-glucose + D-fructose 6-phosphate = UDP +
CC sucrose 6-phosphate.
CC -!- ENZYME REGULATION: ACTIVITY REGULATED BY PHOSPHORYLATION AND
CC MODERATED BY CONCENTRATION OF METABOLITES AND LIGHT.
CC -!- PATHWAY: Sucrose synthesis.
CC -!- SUBUNIT: Homodimer or homotetramer (By similarity).
CC -!- PTM: PHOSPHORYLATED. HOWEVER, PHOSPHORYLATION IS NOT ESSENTIAL FOR
CC ENZYME FUNCTION (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 1.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL; U33175; AAC49379.1; -.
CC PIR; JC4783; JC4783.
CC Gramene; Q43802; -.
CC InterPro; IPR001296; Glyco_transf_1.
CC Pfam; PF00534; Glycosyltransf_1; I.
KW Transferase; Glycosyltransferase; Phosphorylation.
FT DOMAIN 22 29 POLY-GLY.
FT DOMAIN 695 698 POLY-GLU.
FT DOMAIN 775 779 POLY-ARG.
SQ SEQUENCE 1049 AA; 116455 MW; ED862E2819AA4B04 CRC64;

Query Match 66.7%; Score 36; DB 1; Length 1049;
Best Local Similarity 66.7%; Pred. No. 24;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 VVPGXGMDYS 11
   ||| ||:|
Db 436 VVPGXGMDYS 444

RESULT 4
SPS MAIZE
ID SPS MAIZE STANDARD; PRT; 1068 AA.
AC P31927;
DT 01-JUL-1993 (Rel. 26, Created)

```

DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 01-JUN-1994 (Rel. 29, Last annotation update)
 DE Sucrose-phosphate synthase (EC 2.4.1.14) (UDP-glucose-fructose-
 phosphate glucosyltransferase).
 GN SPS.
 OS Zea mays (Maize).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC PACAD clade; Panicoideae; Andropogoneae; Zea.
 ON NCBI_TaxID=4577;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 71-74; 206-212; 471-481 AND
 RP 872-892.
 RC STRAIN=CV. PIONEER 3184; TISSUE=Leaf;
 RX MEDLINE=92338837; PubMed=1840396;
 RA Worrell A.C., Bruneau J.-M., Summerfelt K., Boersig M., Voelker T.A.;
 RA "Expression of a maize sucrose phosphate synthase in tomato alters
 RT leaf carbohydrate partitioning."
 RL Plant Cell 3:1121-1130(1991).
 CC -1- FUNCTION: INVOLVED IN THE REGULATION OF CARBON PARTITIONING IN
 CC THE LEAVES OF PLANTS. MAY REGULATE THE SYNTHESIS OF SUCROSE AND
 CC THEREFORE PLAY A MAJOR ROLE AS A LIMITING FACTOR IN THE EXPORT OF
 CC PHOTOASSIMILATES OUT OF THE LEAF.
 CC -1- CATALYTIC ACTIVITY: UDP-glucose + D-fructose 6-phosphate = UDP +
 CC sucrose 6-phosphate.
 CC -1- ENZYME REGULATION: ACTIVITY REGULATED BY PHOSPHORYLATION AND
 CC MODERATED BY CONCENTRATION OF METABOLITES AND LIGHT.
 CC -1- PATHWAY: Sucrose synthesis.
 CC -1- SUBUNIT: HOMODIMER OR HOMOTETRAMER.
 CC -1- DEVELOPMENTAL STAGE: GERMINATING SEEDS OR MATURE LEAVES.
 CC -1- PTM: PHOSPHORYLATED. HOWEVER, PHOSPHORYLATION IS NOT ESSENTIAL FOR
 CC ENZYME FUNCTION.
 CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 1.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M97550; AAA33513.1; -
 DR PIR; JQ1329; JQ1329.
 DR MaizeDB; 25294; -
 DR InterPro; IPR001296; Glyco_transf_1.
 DR Pfam; PF00534; Glycosyltransferase; Phosphorylation.
 KW Transferase; Glycosyltransferase; Phosphorylation.
 FT DOMAIN 25 31 POLY-GLY.
 FT DOMAIN 25 31
 SQ SEQUENCE 1068 AA; 118575 MW; 074679B5E9A1D282 CRC64;

 Query Match 66.7%; Score 36; DB 1; Length 1068;
 Best Local Similarity 66.7%; Pred. No. 24;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

 Qy 3 VVPXGMDYS 11
 Db 435 VVPXGMDYS 443

 RESULT 5
 SPS2 CRAPL ID_SPS2 CRAPL STANDARD; PRT; 1081 AA.
 AC OA4933;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Sucrose-phosphate synthase 2 (EC 2.4.1.14) (UDP-glucose-fructose-
 phosphate glucosyltransferase 2).
 GN SPS2.
 OS Craterostigma plantagineum.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

OC Asteridae; lamids; Lamiales; Lamiales incertae sedis; Toreneae;
 OC Craterostigma.
 ON NCBI_TaxID=4153;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97451773; PubMed=9306694;
 RA Ingram J., Chandler J.W., Gallagher L., Salami F., Bartels D.;
 RA "Analysis of cDNA clones encoding sucrose-phosphate synthase in
 RT relation to sugar interconversions associated with dehydration in the
 RT resurrection plant *Craterostigma plantagineum* Hochst.";
 RL Plant Physiol. 115:113-121(1997).
 CC -1- FUNCTION: INVOLVED IN THE REGULATION OF CARBON PARTITIONING IN
 CC THE LEAVES OF PLANTS. MAY REGULATE THE SYNTHESIS OF SUCROSE AND
 CC THEREFORE PLAY A MAJOR ROLE AS A LIMITING FACTOR IN THE EXPORT OF
 CC PHOTOASSIMILATES OUT OF THE LEAF.
 CC -1- CATALYTIC ACTIVITY: UDP-glucose + D-fructose 6-phosphate = UDP +
 CC sucrose 6-phosphate.
 CC -1- ENZYME REGULATION: ACTIVITY REGULATED BY PHOSPHORYLATION AND
 CC MODERATED BY CONCENTRATION OF METABOLITES AND LIGHT.
 CC -1- PATHWAY: Sucrose synthesis.
 CC -1- SUBUNIT: Homodimer or homotetramer (By similarity).
 CC -1- PTM: PHOSPHORYLATED. HOWEVER, PHOSPHORYLATION IS NOT ESSENTIAL FOR
 CC ENZYME FUNCTION (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 1.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; Y11795; CAA72491.1; -
 DR PIR; T09837; T09837.
 DR InterPro; IPR001296; Glyco_transf_1.
 DR Pfam; PF00534; Glycosyltransferase; Phosphorylation; Multigene family.
 KW Transferase; Glycosyltransferase; Phosphorylation; Multigene family.
 FT DOMAIN 245 248 POLY-SER.
 FT DOMAIN 256 264 POLY-GLU.
 FT DOMAIN 787 790 POLY-ARG.
 SQ SEQUENCE 1081 AA; 120933 MW; DD142DC2F1A72900 CRC64;

 Query Match 66.7%; Score 36; DB 1; Length 1081;
 Best Local Similarity 66.7%; Pred. No. 24;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

 Qy 3 VVPXGMDYS 11
 Db 445 VVPXGMDYS 453

 RESULT 6
 FAF_DROME ID_FAF_DROME STANDARD; PRT; 2778 AA.
 AC P55824; Q9V9T6; Q9Y027;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable ubiquitin carboxyl-terminal hydrolase FAF (EC 3.1.2.15)
 DE (Ubiquitin thiolesterase FAF) (Ubiquitin-specific processing protease
 DE FAF) (Deubiquitinating enzyme FAF) (Fat facets protein).
 GN FAF OR BCDNA:LD22582 OR CG1945.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 ON NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3), AND TISSUE SPECIFICITY.
 RC TISSUE=Eye imaginal disk;
 RX MEDLINE=93202020; PubMed=1295747;
 RA Fischer-Vize J.A., Rubin G.M., Lehmann R.;


```

ID KHL1 HUMAN STANDARD; PRT; 748 AA.
AC Q9NR64; Q9H4X4; Q9NR65; Q9P238;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Kelch-like protein 1.
GN KHL1 OR KIAA1490.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20347694; PubMed=10888605;
RA Koob M.D., Nemes J.P., Benrow K.A.;
RT "The SCA8 transcript is an antisense RNA to a brain-specific
transcript encoding a novel actin-binding protein (KHL1).";
RL Hum. Mol. Genet. 9:1543-1551(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Nagase T., Kikuno R., Ishikawa K.-I., Hirosewa M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XVII.
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro.";
RL DNA Res. 7:143-150(2000).
RN [3]
RP SEQUENCE OF 179-409 FROM N.A.
RA Kay M.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAY PLAY A ROLE IN ORGANIZING THE ACTIN CYTOSKELETON OF
THE BRAIN CELLS.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: Highly expressed in brain.
CC -1- SIMILARITY: Contains 1 BTB/POZ domain.
CC -1- SIMILARITY: Contains 6 Kelch repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
EMBL; AF252283; AAP81719.1; -.
EMBL; AF252279; AAF81716.1; -.
EMBL; AB040923; BAA36014.1; ALT_INIT.
EMBL; AL353738; CAC16128.1; -.
GENE; HGNC:6352; KHL1.
MIM; 605332;
GO; GO:0003779; P:actin binding activity; NAS.
GO; GO:0030036; P:actin cytoskeleton organization and biogenesis; NAS.
InterPro; IPR000210; BTB_POZ.
InterPro; IPR006652; Kelch_rep.
Pfam; PF00651; BTB; 1.
Pfam; PF01344; Kelch; 6.
SMART; SM00225; BTB; 1.
SMART; SM00612; Kelch; 6.
PROSITE; PS00097; BTB; 1.
CytoSkeleto; Actin-binding; Kelch repeat; Repeat.
DOMAIN 43 88
DOMAIN 212 279
REPEAT 460 506
REPEAT 507 553
REPEAT 555 600
REPEAT 601 647
REPEAT 649 700
REPEAT 701 747
REPEAT 748 AA; 82680 MW; C11C43D8282P9FF9 CRC64;
SQ SEQUENCE

```

Query Match

63.9%; Score 34.5; DB 1; Length 748;

```

Best Local Similarity 80.0%; Pred. No. 34;
Matches 8; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

QY 1 BEVVPXGWDY 10
DB 127 BEVVP-GMDP 135

RESULT 8
PLAS ORYSA STANDARD; PRT; 154 AA.
AC P20423; Q9SBB8;
DT 01-FEB-1991 (Rel. 17, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Plastocyanin, chloroplast precursor.
GN PTEB.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OC NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Ilpoom; TISSUE=Leaf;
RA Lee J.-S.;
RT "Molecular cloning and characterization of plastocyanin precursor in
rice.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 58-154.
RC STRAIN=cv. Japonica;
RX MEDLINE=89386623; PubMed=2780537;
RA Yano H., Kano M., Tsugita A., Aso K., Nozu Y.;
RT "The amino acid sequence of plastocyanin from rice (Oryza sativa,
subspecies japonica).";
RL Protein Seq. Data Anal. 2:385-389(1989).
CC -1- FUNCTION: Participates in electron transfer between P700 and the
cytochrome b6-f complex in photosystem I.
CC -1- SUBCELLULAR LOCATION: LOOSELY BOUND TO THE INNER THYLAKOID
MEMBRANE SURFACE. Contains 1 plastocyanin-like domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
EMBL; AF093636; AAC78108.1; -.
HSP; P00289; 2PCF.
Gramene; P20423; -.
InterPro; IPR000923; BlueCu_1.
InterPro; IPR001235; Copper_blue.
Pfam; PF00127; copper_bind; 1.
PRINTS; PR00156; COPPERBLUE.
ProDom; PD001235; Copper_blue; 1.
PROSITE; PS00196; COPPER_BLUE; 1.
KW Chloroplast; Electron transport; Copper; Thylakoid; Membrane;
KW Transit peptide.
FT TRANSIT 1 57 CHLOROPLAST.
FT CHAIN 58 154 PLASTOCYANIN.
FT DOMAIN 58 154 PLASTOCYANIN-LIKE.
FT METAL 94 94 COPPER (BY SIMILARITY).
FT METAL 139 139 COPPER (BY SIMILARITY).
FT METAL 142 142 COPPER (BY SIMILARITY).
FT METAL 147 147 COPPER (BY SIMILARITY).
SQ SEQUENCE 154 AA; 15577 MW; E45725D25B5P400D CRC64;

```

Query Match

63.0%; Score 34; DB 1; Length 154;
Best Local Similarity 54.5%; Pred. No. 8.1;

Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 REVVPXGMDYS 11
|: || |:
Db 100 EDAPVSGVDVS 110

RESULT 9

ID PLAS_HORVU STANDARD; PRT; 155 AA.
AC P08248;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Plastocyanin, chloroplast precursor.
GN PETE.
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
OC Triticeae; Hordeum.
OC eurosida I; Malpighiales; Salicaceae; Populus.
OX NCBI_TaxID=4513;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Bomi;
RA Nielsen O.S., Gausing K.;
RT "The precursor of barley plastocyanin: sequence of cDNA clones and
RT gene expression in different tissues.";
RL FEBS Lett. 225:159-162(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. NK 1558;
RX MEDLINE=94039081; PubMed=8223592;
RA Nielsen P., Gausing K.;
RT "In vitro binding of nuclear proteins to the barley plastocyanin gene
RT promoter region.";
RL Eur. J. Biochem. 217:97-104(1993).
CC -!- FUNCTION: Participates in electron transfer between P700 and the
CC cytochrome b6-f complex in photosystem I.
CC -!- SUBCELLULAR LOCATION: LOOSELY BOUND TO THE INNER THYLAKOID
CC MEMBRANE SURFACE IN CHLOROPLASTS.
CC -!- SIMILARITY: Contains 1 plastocyanin-like domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC EMBL; Y00704; CAA68696.1; -
DR EMBL; Z28347; CAA82201.1; -
DR PIR; S38255; S38255.
DR HSSP; P00289; 2PCF.
DR InterPro; IPR000923; BlueCu 1.
DR Pfam; PF00127; copper-bind; 1.
DR PRINTS; PR00156; COPPERBLUE.
DR PRODOM; PD001235; Copper blue; 1.
DR PROSITE; PS00196; COPPER BLUE; 1.
DR Chloroplast; Electron transport; Copper; Thylakoid; Membrane;
KW Transit peptide.
FT TRANSIT 1 58 CHLOROPLAST.
FT CHAIN 59 155 PLASTOCYANIN.
FT DOMAIN 59 155 PLASTOCYANIN-LIKE.
FT METAL 59 95 COPPER (BY SIMILARITY).
FT METAL 140 140 COPPER (BY SIMILARITY).
FT METAL 143 143 COPPER (BY SIMILARITY).
FT METAL 148 148 COPPER (BY SIMILARITY).
FT VARIANT 120 120 T -> N (IN CV. NK 1558).
SQ SEQUENCE 155 AA; 15709 MW; DAA7EABE5F6F4F91 CRC64;

Query Match 63.0%; Score 34; DB 1; Length 155;

Best Local Similarity 54.5%; Pred. No. 8.2;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 REVVPXGMDYS 11
|: || |:
Db 101 EDAPVSGVDVS 111

RESULT 10

ID PLAT_POPNI STANDARD; PRT; 168 AA.
AC P11970;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Plastocyanin B, chloroplast precursor.
GN PETE.
OS Populus nigra (Lombardy poplar).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosida I; Malpighiales; Salicaceae; Populus.
OX NCBI_TaxID=3691;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Italica; TISSUE=Leaf;
RA Reichert J., Jenzelewski V., Haehnel W.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 70-168.
RC STRAIN=cv. Italica;
RA Dimitrov M.I., Egorov C.A., Donchev A.A., Atanasov B.P.;
RT "Complete amino acid sequence of poplar plastocyanin b.";
RL FEBS Lett. 226:17-22(1987).
CC -!- FUNCTION: Participates in electron transfer between P700 and the
CC cytochrome b6-f complex in photosystem I.
CC -!- SUBCELLULAR LOCATION: LOOSELY BOUND TO THE INNER THYLAKOID
CC MEMBRANE SURFACE IN CHLOROPLASTS.
CC -!- SIMILARITY: THERE ARE 12 DIFFERENCES BETWEEN THE SEQUENCE OF
CC POPLAR PLASTOCYANINS A AND B.
CC -!- SIMILARITY: Contains 1 plastocyanin-like domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC EMBL; Z50186; CAA90565.1; -
DR PIR; S00210; S00210.
DR PIR; S58208; S58208.
DR HSSP; P00299; 1PLC.
DR InterPro; IPR000923; BlueCu 1.
DR InterPro; IPR001235; Copper blue.
DR Pfam; PF00127; copper-bind; 1.
DR PRINTS; PR00156; COPPERBLUE.
DR PRODOM; PD001235; Copper blue; 1.
DR PROSITE; PS00196; COPPER BLUE; 1.
DR Chloroplast; Electron transport; Copper; Thylakoid; Membrane;
KW Transit peptide; Multigene family.
FT TRANSIT 1 69 CHLOROPLAST.
FT CHAIN 70 168 PLASTOCYANIN B.
FT DOMAIN 70 168 PLASTOCYANIN-LIKE.
FT METAL 106 106 COPPER.
FT METAL 153 153 COPPER.
FT METAL 156 156 COPPER.
FT METAL 161 161 COPPER.
SQ SEQUENCE 168 AA; 16981 MW; F20DA6EA2038AEEA CRC64;

Query Match 63.0%; Score 34; DB 1; Length 168;
Best Local Similarity 54.5%; Pred. No. 8.9;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVFXGMDYS 11
: ||| :
DB 112 EDVPSGVDVS 122

RESULT 11

ID SYFB BORBU STANDARD; PRT; 566 AA.
AC P94283;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phenylalanyl-tRNA synthetase beta chain (EC 6.1.1.20) (Phenylalanine--
DE tRNA ligase beta chain) (PHERS).
GN PHER OR BB0514.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RA Barbour A.G.; Himebusch J.;
RT "Phenylalanyl-tRNA synthetase genes (alpha and beta subunits) and
RT thioesterase gene of Borrelia burgdorferi.";
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA Dougherty B., Tomb J.-P., Fleischmann R.D., Richardson D.,
RA Peterson J., Kerlavage A.R., Quackenbush J.J., Salzberg S., Hanson M.,
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA Uterback T., Wathley L., McDonald L., Artach P., Bowman C.,
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochete, Borrelia
RT burgdorferi";
RL Nature 390:580-586(1997).
CC -1- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA (Phe) = AMP +
CC diphosphate + L-phenylalanyl-tRNA (Phe).
CC -1- SUBUNIT: Tetramer of two alpha and two beta chains (by
CC similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE PHENYLALANYL-TRNA SYNTHETASE BETA
CC CHAIN FAMILY. SUBFAMILY 2.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; U82978; AB41019.1; -
EMBL; AB001153; AAC66870.1; -
DR PIR; A70164; A70164.
DR HAMAP; BB0514; -
DR InterPro; IPR005147; B5.
DR InterPro; IPR004531; PHER_arch.
DR Pfam; PF03484; B5; 1.
DR TIGRFAMs; TIGR00471; pHER_arch; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
SQ SEQUENCE 566 AA; 65173 MW; 9D48C8B5D6D3B74B CRC64;

Query Match 63.0%; Score 34; DB 1; Length 566;
Best Local Similarity 85.7%; Pred. No. 32;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 VPXGMDY 10
: ||| :
DB 169 VPGMDY 175

RESULT 12

ID RBL4 ARATH STANDARD; PRT; 908 AA.
AC Q9FK9;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable disease resistance RPP8-like protein 4.
GN RPP8L4 OR AT5G48620 OR K15N18.9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=99087489; PubMed=9872454;
RA Nakamura Y., Sato S., Asamizu E., Kaneko T., Kotani H., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. VII.
RT Sequence features of the regions of 1,013,767 bp covered by sixteen
RT physically assigned P1 and TAC clones.";
RL DNA Res. 5:297-308(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,
RA Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,
RA Hayashizaki Y., Shinozaki K.;
RT "Arabidopsis thaliana full-length cDNA.";
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Potential disease resistance protein.
CC -1- DOMAIN: The LRR repeats probably act as specificity determinant of
CC pathogen recognition (by similarity).
CC -1- SIMILARITY: Belongs to the disease resistance NB-LRR family.
CC RPP8/LRR subfamily.
CC -1- SIMILARITY: Contains 3 leucine-rich (LRR) repeats.
CC -1- SIMILARITY: Contains 1 NB-ARC domain.
CC -1- DATABASE: NAMS-NIB-LRRS;
CC NOTES=Functional and comparative genomics of disease resistance gene
CC homologs;
CC WWW=<http://niblr.ucdavis.edu>.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AB015468; BAB10695.1; -
EMBL; AK117163; BAC41841.1; -
DR InterPro; IPR000767; Disease_resist.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR002182; NB-ARC.
DR Pfam; PF00560; LRR; 2.
DR Pfam; PF00931; NB-ARC; 1.
DR PRINTS; PR00364; DISSEASERSIST.
KW Plant defense; ATP-binding; Repeat; Leucine-rich repeat.
FT DOMAIN 10 45 LEUCINE-ZIPPER.
FT DOMAIN 146 459 NB-ARC.
FT REPEAT 575 599 LRR 1.
FT REPEAT 600 623 LRR 2.
FT REPEAT 842 867 LRR 3.
FT NP_BIND 192 199 ATP (POTENTIAL).

| | | |
|-----------|---|--|
| SQ | SEQUENCE | 908 AA; 104448 MW; 3111991B1723693 CRC64; |
| | Query Match | 63.0%; Score 34; DB 1; Length 908; |
| | Best Local Similarity | 60.0%; Pred. No. 53; |
| | Matches | 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0; |
| OY | 1 BEVPXGMDY 10 :: | |
| Db | 883 EKLVPGEDEY 892 | |
| RESULT 13 | | |
| ID | RPP8_ARATH | PRT; 908 AA. |
| AC | Q8WJ79; Q8GWG5; Q9MSA1; Q9ZSY3; Q9ZY4; | |
| DT | 15-SEP-2003 (Rel. 42, Created) | |
| DT | 15-SEP-2003 (Rel. 42, Last sequence update) | |
| DT | 15-SEP-2003 (Rel. 42, Last annotation update) | |
| DE | Disease resistance protein RPP8 (Resistance to Peronospora parasitica protein 8). | |
| GN | RPP8 OR HRT OR AT5G43470 OR MWF20.19. | |
| OS | Arabidopsis thaliana (Mouse-ear cress). | |
| OC | Eukaryota; Viridiplantae; Streptophyta; Tracheophyta; | |
| OC | Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae; | |
| OC | eurosid II; Brassicales; Brassicaceae; Arabidopsida. | |
| ON | NCBI_TaxID=3702; | |
| OR | [1] | |
| RP | SEQUENCE FROM N.A. (ISOFORM 1), FUNCTION, MUTANTS RPP8-1; RPP8-2 AND RPP8-3, AND VARIANTS. | |
| RC | STRAIN=cv. Columbia, and cv. Landsberg erecta; | |
| RX | MEDLINE=99030193; PubMed=9811794; | |
| RA | McDowell J.M., Dhandaydham M., Long T.A., Aarts M.G.M., Goff S., Holub E.B., Dangl J.L.; | |
| RT | "Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the RPP8 locus of Arabidopsis."; | |
| RL | Plant Cell 10:1861-1874 (1998) . | |
| RN | (2) | |
| RP | SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANTS. | |
| RC | STRAIN=cv. Di-17; | |
| RX | MEDLINE=2071766; PubMed=10810142; | |
| RA | Cooley M.B., Fathirana S., Wu H.-J., Kachroo P., Kleissig D.F.; | |
| RT | "Members of the Arabidopsis HRT/RPP8 family of resistance genes confer resistance to both viral and oomycete pathogens."; | |
| RL | Plant Cell 12:663-676(2000) . | |
| RN | (3) | |
| RP | SEQUENCE FROM N.A. (ISOFORM 1). | |
| RC | STRAIN=cv. Columbia; | |
| RX | MEDLINE=20181125; PubMed=10718197; | |
| RA | Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H., Tabata S.; | |
| RT | "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence features of the regions of 3,076,755 bp covered by sixty Pl and TAC clones."; | |
| RL | DNA Res. 7:31-63(2000) . | |
| RN | (4) | |
| RP | SEQUENCE FROM N.A. (ISOFORM 1). | |
| RC | STRAIN=cv. Columbia; | |
| RA | Shinozaki K., Davis R.W., Ecker J.R., Theologis A.; | |
| RT | "Riken Arabidopsis full length cDNA clones (RAFTs) sequenced by the SSP consortium (Salk/Stanford/PGECC)."; | |
| RL | Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases. | |
| RN | (5) | |
| RP | SEQUENCE FROM N.A. (ISOFORM 2). | |
| RC | STRAIN=cv. Columbia; | |
| RA | Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J., Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J., Hayashizaki Y., Shinozaki K.; | |
| RT | "Arabidopsis thaliana full-length cDNA."; | |
| RL | Submitted (NOV-2002) to the EMBL/GenBank/DDBJ databases. | |
| RN | (6) | |
| RP | INTERACTION WITH TIP. | |
| RT | MEDLINE=20496823; PubMed=11041886; | |
| RX | | |

| | | |
|----|---|--|
| RA | Ren T., Qu F., Morris T.J.; | This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.ebi.ac.uk/annotation/or send an email to license@ebi.ac.uk). |
| RT | *HRT gene function requires interaction between a NAC protein and viral capsid protein to confer resistance to turnip crinkle virus." | |
| RL | Plant Cell 12:1917-1926(2000). | |
| CC | -!- FUNCTION: Disease resistance protein. Resistance proteins guard the plant against pathogens that contain an appropriate avirulence protein via an indirect interaction with this avirulence protein. That triggers a defense system including the hypersensitive response, which restricts the pathogen growth. The interaction with TIP (TCV-interacting protein) may be essential for the recognition of the avirulence proteins, and the triggering of the defense response. | |
| CC | -!- SUBUNIT: Interacts with the NAC protein TIP. | |
| CC | -!- ALTERNATIVE PRODUCTS: | |
| CC | Event=Alternative splicing; Named isoforms=2; | |
| CC | Name=1; | |
| CC | IsoId=Q8W4J9-1; Sequence=Displayed; | |
| CC | Name=2; | |
| CC | IsoId=Q8W4J9-2; Sequence=VSP_007171, VSP_007172; | |
| CC | Note=Has been shown to exist only in cv. Columbia so far; | |
| CC | -!- DOMAIN: The LRR repeats probably act as specificity determinant of pathogen recognition. | |
| CC | -!- POLYMORPHISM: The strong polymorphisms present in cv. Di-17 and cv. Columbia are probably due to an unequal crossing-over between the highly related RPP8 and RPP8A genes present in cv. Landsberg erecta. Such variations probably modify the specificity of pathogen recognition. | |
| CC | -!- MISCELLANEOUS: In cv. Columbia and cv. Landsberg erecta, RPP8 specifically recognizes the Emco5 avirulence protein from Peronospora parasitica, while it is not the case in cv. Di-17, where it confers resistance to turnip crinkle virus upon recognition of the viral capsid protein. | |
| CC | -!- SIMILARITY: Belongs to the disease resistance NB-LRR family. RPP8/HRT subfamily. | |
| CC | -!- SIMILARITY: Contains 2 leucine-rich (LRR) repeats. | |
| CC | -!- SIMILARITY: Contains 1 NB-ARC domain. | |
| CC | -!- DATABASE: NAME=NIB-LRRS; | |
| CC | NOTE=Functional and comparative genomics of disease resistance gene homologs; | |
| CC | WWW="http://niblrns.ucdavis.edu". | |
| DR | EMBL; AF089710; AAC83165.1; - | |
| DR | EMBL; AF089711; AAC78631.1; - | |
| DR | EMBL; AF234174; AAF36987.1; - | |
| DR | EMBL; AB025638; BAA37436.1; - | |
| DR | EMBL; AF062514; AAL32592.1; - | |
| DR | EMBL; AKI18862; BAC43449.1; - | |
| DR | InterPro; IPR000767; Disease_resist. | |
| DR | InterPro; IPR001611; LRR. | |
| DR | InterPro; IPR002182; NB-ARC. | |
| DR | Pfam; PF00560; LRR; 2. | |
| DR | Pfam; PF00931; NB-ARC; 1. | |
| DR | PRINTS; PR00364; DISEASERSIST. | |
| KW | Plant defense; ATP-binding; Repeat; Leucine-rich repeat; Alternative splicing; Polymorphism. | |
| FT | DOMAIN 10 45 LEUCINE-ZIPPER. | |
| FT | DOMAIN 146 459 NB-ARC. | |
| FT | REPEAT 600 623 LRR 1. | |
| FT | REPEAT 842 867 LRR 2. | |
| FT | NP BIND 192 199 ATP (POTENTIAL). | |
| FT | VARSPLIC 294 308 WKLLTSRNEGVIH -> ELLMYHEALFLNS (in isoform 2) | |
| FT | /FTrid=VSP_007171. | |
| FT | Missing (in isoform 2). | |
| FT | /FTTd=VSP_007172 | |
| FT | VARSPLIC 309 908 | |

[illegible]

FT VARIANT 27 31 IDQL -> VDEQI (in cv. Landsberg erecta).
 FT VARIANT 29 29 G -> E (in cv. Di-17).
 FT VARIANT 85 87 SGK -> RGE (in cv. Di-17 and cv. Landsberg erecta).
 FT VARIANT 88 91 GKGV -> BKG1 (in cv. Landsberg erecta).
 FT VARIANT 100 100 C -> R (in cv. Landsberg erecta).
 FT VARIANT 121 121 E -> D (in cv. Landsberg erecta).
 FT VARIANT 129 129 F -> L (in cv. Di-17).
 FT VARIANT 133 133 Q -> Q (in cv. Di-17).
 FT VARIANT 138 138 G -> V (in cv. Landsberg erecta).
 FT VARIANT 173 173 K -> T (in cv. Di-17).
 FT VARIANT 177 177 G -> C (in cv. Di-17).
 FT VARIANT 253 253 Y -> S (in cv. Di-17).
 FT VARIANT 257 257 R -> P (in cv. Di-17).
 FT VARIANT 265 265 A -> T (in cv. Di-17).
 FT VARIANT 267 267 R -> K (in cv. Landsberg erecta).
 FT VARIANT 270 270 V -> L (in cv. Di-17).
 FT VARIANT 283 283 V -> R (in cv. Di-17).
 FT VARIANT 369 369 A -> V (in cv. Di-17 and cv. Landsberg erecta).
 FT VARIANT 387 387 F -> S (in cv. Di-17 and cv. Landsberg erecta).
 FT VARIANT 399 399 W -> C (in cv. Di-17 and cv. Landsberg erecta).
 FT VARIANT 426 426 C -> R (in cv. Di-17 and cv. Landsberg erecta).
 FT VARIANT 429 429 N -> F (in cv. Di-17 and cv. Landsberg erecta).
 FT VARIANT 436 446 DSEISTYSLFY -> YKISAYDLFN (in cv. Landsberg erecta).
 FT VARIANT 438 443 EISTYS -> KITQE (in cv. Di-17).
 FT VARIANT 450 450 A -> V (in cv. Landsberg erecta).
 FT VARIANT 460 460 E -> Q (in cv. Landsberg erecta).
 FT VARIANT 480 486 DNYLSWQ -> NKYLVRH (in cv. Di-17).
 FT VARIANT 480 481 DN -> NR (in cv. Landsberg erecta).
 FT VARIANT 485 486 WQ -> SH (in cv. Landsberg erecta).
 FT VARIANT 489 489 Y -> N (in cv. Landsberg erecta).
 FT VARIANT 514 514 I -> K (in cv. Di-17 and cv. Landsberg erecta).
 FT VARIANT 518 518 C -> S (in cv. Di-17 and cv. Landsberg erecta).
 FT VARIANT 519 519 T -> I (in cv. Di-17).
 FT VARIANT 528 528 S -> R (in cv. Di-17).
 FT VARIANT 537 537 S -> R (in cv. Di-17).
 FT VARIANT 542 542 H -> Q (in cv. Di-17).
 FT VARIANT 543 543 I -> L (in cv. Landsberg erecta).
 FT VARIANT 547 550 KNKT -> RNAA (in cv. Di-17).
 FT VARIANT 549 549 K -> N (in cv. Landsberg erecta).
 FT VARIANT 558 565 PRFEDYD -> WDEDFG (in cv. Landsberg erecta).
 FT VARIANT 558 560 PRF -> SRFK (in cv. Di-17).
 FT VARIANT 564 564 Y -> F (in cv. Di-17).
 FT VARIANT 584 584 S -> Y (in cv. Landsberg erecta).
 FT VARIANT 595 595 C -> S (in cv. Di-17 and cv. Landsberg erecta).
 FT VARIANT 609 612 YEAK -> FLAG (in cv. Landsberg erecta).
 FT VARIANT 630 635 NLVDT -> DLSVHE (in cv. Di-17).
 FT VARIANT 632 636 RVDTE -> SVNNK (in cv. Landsberg erecta).
 FT VARIANT 650 650 Q -> E (in cv. Di-17).
 FT VARIANT 653 654 YL -> HI (in cv. Di-17).
 FT VARIANT 659 659 Missing (in cv. Di-17).
 FT VARIANT 660 660 M -> K (in cv. Landsberg erecta).
 FT VARIANT 676 678 YLY -> PLP (in cv. Landsberg erecta).
 FT VARIANT 678 679 YG -> FR (in cv. Di-17).
 FT VARIANT 692 692 R -> H (in cv. Landsberg erecta).
 FT VARIANT 697 697 R -> Q (in cv. Di-17).
 FT VARIANT 700 700 A -> G (in cv. Di-17).

Query Match 63.0%; Score 34; DB 1; Length 908;
 Best Local Similarity 60.0%; Pred. No. 53;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 BEVVPXGMDY 10

Db 883 EKLVPGGEDY 892

RESULT 14

RP8H ARATH
 ID RP8H ARATH STANDARD; PRT; 910 AA.
 AC P59584;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Disease resistance protein RPH8A (RPP8 homolog A).
 GN RPH8A.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A., AND FUNCTION.
 RC STRAIN=cv. Landsberg erecta;
 RX MEDLINE=99030193; PubMed=9811794;
 RA McDowell J.M., Dhandaydham M., Long T.A., Aarts M.G.M., Goff S.,
 RA Holub E.B., Dangl J.L.;
 RT "Intragenic recombination and diversifying selection contribute to the
 evolution of downy mildew resistance at the RPP8 locus of
 Arabidopsis.";
 RL Plant Cell 10:1861-1874(1998).
 CC -1- FUNCTION: Disease resistance protein. Resistance proteins guard
 the plant against pathogens that contain an appropriate avirulence
 protein via an indirect interaction with this avirulence protein.
 CC That triggers a defense system including the hypersensitive
 response, which restricts the pathogen growth. In contrast to
 RPP8, it does not specifically recognize the Emco5 avirulence
 protein from Peronospora parasitica.
 CC -1- MISCELLANEOUS: In cv. Columbia and cv. Di-17, this protein is not
 present due to an unequal crossing over between the RPP8 and RPH8A
 genes that creates a unique RPP8 gene.
 CC -1- SIMILARITY: Belongs to the disease resistance NB-LRR family.
 CC RPP8/HRT subfamily.
 CC -1- SIMILARITY: Contains 2 leucine-rich (LRR) repeats.
 CC -1- SIMILARITY: Contains 1 NB-ARC domain.
 CC -1- DATABASE: NAME=NIB-LRRS;
 CC NOTE=Functional and comparative genomics of disease resistance gene
 homologs;
 CC WWW=http://niblr.ucdavis.edu".
 KW Plant defense; ATP-binding; Repeat; Leucine-rich repeat.
 FT DOMAIN 10 45 LEUCINE-ZIPPER.
 FT DOMAIN 146 459 NB-ARC.
 FT REPEAT 602 625 LRR 1.
 FT REPEAT 844 869 LRR 2.
 FT NP BIND 192 199 ATP (POTENTIAL).
 SQ SEQUENCE 910 AA; 105263 MW; 5B189F65A19A12EE CRC64;
 Query Match 63.0%; Score 34; DB 1; Length 910;
 Best Local Similarity 60.0%; Pred. No. 53;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 BEVVPXGMDY 10
 Db 885 EKLVPGGEDY 894

RESULT 15
 Y939 METJA
 ID Y939 METJA STANDARD; PRT; 276 AA.
 AC Q58349;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein MJ0939.
 GN MJ0939.
 OS Methanococcus jannaschii.

Tue Dec 23 09:51:04 2003

```

OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kervatage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness B.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RL jannaschii."
PL Science 273:1058-1073(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; U67537; AAB98946.1; -.
DR PIR: C64417; C64417.
DR TIGR: MJ0939; -.
DR Pfam: PF03692; UPP0153; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 276 AA; 33454 MW; 97BD69D392BC8FDF CRC64;

Query Match 61.1%; Score 33; DB 1; Length 276;
Best Local Similarity 45.5%; Pred. No. 24;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EEVVPXGMDYS 11
Db 141 EEIENGWEHS 151

```

Search completed: December 22, 2003, 17:42:27
 Job time : 5.6 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 17:27:26 ; Search time 25.2 Seconds
(without alignments)
112.642 Million cell updates/sec

Title: US-09-909-164-13

Perfect score: 54

Sequence: 1 EVVVPXGMDYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23.*

- 1: sp_archea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriap.*
- 17: sp_archeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-----------|---------------------|
| 1 | 40 | 74.1 | 156 | 3 Q12479 | Q12479 saccharomyc |
| 2 | 38 | 70.4 | 319 | 16 Q8ESV7 | Q8ESV7 oceanobacill |
| 3 | 38 | 70.4 | 363 | 17 Q30260 | Q30260 archaeoglob |
| 4 | 38 | 70.4 | 1044 | 16 Q8DIH0 | Q8DIH0 synchococcc |
| 5 | 36 | 66.7 | 341 | 10 Q22081 | Q22081 citrus uneh |
| 6 | 36 | 66.7 | 348 | 10 Q22096 | Q22096 citrus uneh |
| 7 | 36 | 66.7 | 452 | 10 Q8W568 | Q8W568 arabidopsis |
| 8 | 36 | 66.7 | 460 | 10 Q9C9T7 | Q9C9T7 arabidopsis |
| 9 | 36 | 66.7 | 1047 | 10 P93782 | P93782 saccharum o |
| 10 | 36 | 66.7 | 1083 | 10 Q9SN30 | Q9SN30 arabidopsis |
| 11 | 36 | 66.7 | 1084 | 10 Q43010 | Q43010 oryza sativ |
| 12 | 36 | 66.7 | 1100 | 10 Q8S064 | Q8S064 oryza sativ |
| 13 | 35 | 64.8 | 219 | 5 Q9GQ04 | Q9GQ04 eriocheir s |
| 14 | 35 | 64.8 | 253 | 16 Q8XPA8 | Q8XPA8 clostridium |
| 15 | 35 | 64.8 | 298 | 2 Q52367 | Q52367 rhizobium t |
| 16 | 35 | 64.8 | 368 | 2 Q8KL43 | Q8KL43 rhizobium e |

| | | | | | |
|----|------|------|------|-----------|--------------------|
| 17 | 35 | 64.8 | 425 | 5 Q9XVK4 | Q9XVK4 caenorhabdi |
| 18 | 35 | 64.8 | 433 | 16 Q9A382 | Q9A382 caulobacter |
| 19 | 35 | 64.8 | 440 | 17 Q9YF13 | Q9YF13 aeropyrum p |
| 20 | 35 | 64.8 | 511 | 2 Q52680 | Q52680 escherichia |
| 21 | 35 | 64.8 | 517 | 16 Q8XZL5 | Q8XZL5 raietonia s |
| 22 | 35 | 64.8 | 595 | 10 Q946J7 | Q946J7 androgaphi |
| 23 | 35 | 64.8 | 745 | 5 Q9SP46 | Q9SP46 carinus ma |
| 24 | 35 | 64.8 | 1031 | 5 Q9UEA3 | Q9UEA3 callinectes |
| 25 | 35 | 64.8 | 1150 | 5 Q17704 | Q17704 caenorhabdi |
| 26 | 35 | 64.8 | 1410 | 2 Q52673 | Q52673 escherichia |
| 27 | 35 | 64.8 | 1420 | 2 Q52666 | Q52666 escherichia |
| 28 | 35 | 64.8 | 1474 | 17 Q27146 | Q27146 methanobact |
| 29 | 35 | 64.8 | 1828 | 16 Q98K29 | Q98K29 rhizobium l |
| 30 | 34.5 | 63.9 | 748 | 4 Q8TBJ7 | Q8TBJ7 homo sapien |
| 31 | 34 | 63.0 | 143 | 5 Q9VSY8 | Q9VSY8 drosophila |
| 32 | 34 | 63.0 | 215 | 16 Q8B9L5 | Q8B9L5 thermoaer |
| 33 | 34 | 63.0 | 222 | 16 Q8E5L9 | Q8E5L9 streptococ |
| 34 | 34 | 63.0 | 282 | 16 Q8DZW9 | Q8DZW9 streptococ |
| 35 | 34 | 63.0 | 284 | 2 Q8GP33 | Q8GP33 lactobacill |
| 36 | 34 | 63.0 | 290 | 16 Q8U7J0 | Q8U7J0 agrobacteri |
| 37 | 34 | 63.0 | 296 | 17 Q9YET8 | Q9YET8 aeropyrum p |
| 38 | 34 | 63.0 | 357 | 17 Q29920 | Q29920 archaeoglob |
| 39 | 34 | 63.0 | 366 | 17 Q29451 | Q29451 archaeoglob |
| 40 | 34 | 63.0 | 387 | 16 Q98FX1 | Q98FX1 rhizobium l |
| 41 | 34 | 63.0 | 543 | 3 Q8TFP4 | Q8TFP4 trichoderma |
| 42 | 34 | 63.0 | 558 | 16 Q8R8Z2 | Q8R8Z2 thermoaer |
| 43 | 34 | 63.0 | 565 | 16 Q9CIN1 | Q9CIN1 lactococcus |
| 44 | 34 | 63.0 | 587 | 16 Q9JZP8 | Q9JZP8 neisseria m |
| 45 | 34 | 63.0 | 671 | 16 Q8G3J2 | Q8G3J2 bifidobacte |

ALIGNMENTS

RESULT 1

| | | | |
|---|--------------|------|---------|
| Q12479 | PRELIMINARY; | PRT; | 156 AA. |
| ID Q12479 | | | |
| AC Q12479; | | | |
| DT 01-NOV-1996 (T-EMBLrel. 01, Created) | | | |
| DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update) | | | |
| DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update) | | | |
| DE ORP YOR013W. | | | |
| GN YOR013W. | | | |
| OS Saccharomyces cerevisiae (Baker's yeast). | | | |
| OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; | | | |
| OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes. | | | |
| OX NCBI_TaxID=4932; | | | |
| RN [1] | | | |
| RP SEQUENCE FROM N.A. | | | |
| RA De haan M., Grivell L.A., Maarse A.C.; | | | |
| RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases. | | | |
| RN [2] | | | |
| RP SEQUENCE FROM N.A. | | | |
| RA MIPS; | | | |
| RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases. | | | |
| RN [3] | | | |
| RP SEQUENCE FROM N.A. | | | |
| RC STRAIN=FY1679; | | | |
| RA De haan M., Maarse A.C., Grivell L.A.; | | | |
| RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases. | | | |
| RN [4] | | | |
| RP SEQUENCE FROM N.A. | | | |
| RC STRAIN=FY1679; | | | |
| RA MEDLINE=94019318; PubMed=8413243; | | | |
| RA Dumont M.E., Schlichter J.B., Cardillo T.S., Hayes M.K., Bethlenny G., | | | |
| RT "CYC2 encodes a factor involved in mitochondrial import of yeast | | | |
| RT cytochrome c."; | | | |
| RL Mol. Cell. Biol. 13:6442-6451(1993). | | | |
| RN [5] | | | |
| RP SEQUENCE FROM N.A. | | | |
| RC STRAIN=FY1679; | | | |
| RA MEDLINE=94169519; PubMed=7764548; | | | |

RA Lee Y.S., Shimizu J., Yoda K., Yamaaki M.;
 RT "Molecular cloning of a gene, DHS1, which complements a drug-
 RT hypersensitive mutation of the yeast *Saccharomyces cerevisiae*.";
 RL Bioeci. Biotechnol. Biochem. 58:391-395(1994).
 DR EMBL; 274920; CAA9201.1; -;
 DR EMBL; X87331; CAA60762.1; -;
 DR SGD; S0005539; YOR013W.
 SQ SEQUENCE 156 AA; 17881 MW; 380442B74C272B41 CRC64;
 Query Match 74.1%; Score 40; DB 3; Length 156;
 Best Local Similarity 77.8%; Pred. No. 2.3; Mismatches 1; Indels 0; Gaps 0;
 Matches 7; Conservative 1;
 Qy 2 EVVPXGMDY 10
 ||:|:|:|
 Db 50 EVMPXGMDY 58
 RESULT 2
 Q8ESV7 PRELIMINARY; PRT; 319 AA.
 AC Q8ESV7;
 DT 01-MAR-2003 (TRENBLrel. 23, Created)
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE Hypothetical conserved protein.
 GN OB0509.
 OS Oceanobacillus thelyensis.
 OC Bacteria; Firmicutes; Bacillales; Oceanobacillus.
 OX NCBI_TaxID=182710;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HTB831 / DSM 14371 / JCM 11309;
 RX MEDLINE=2220767; PubMed=12235376;
 RA Takami H., Takaki Y., Uchiyama I.;
 RT "Genome sequence of Oceanobacillus thelyensis isolated from the Iheya
 RT Ridge and its unexpected adaptive capabilities to extreme
 RT environments.";
 RL Nucleic Acids Res. 30:3927-3935(2002).
 DR EMBL; AP004594; BAC12465.1; -;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 319 AA; 35617 MW; 3BDAE4BF13E79E37 CRC64;
 Query Match 70.4%; Score 38; DB 16; Length 319;
 Best Local Similarity 60.0%; Pred. No. 14;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 BEVVPXGMDY 10
 ||:|:|:|
 Db 189 EQLVPHGIDY 198
 RESULT 3
 O30260 PRELIMINARY; PRT; 363 AA.
 AC O30260;
 DT 01-JAN-1998 (TRENBLrel. 05, Created)
 DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
 DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
 DE Hypothetical protein AF2411.
 GN AF2411.
 OS Archaeoglobus fulgidus.
 OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
 OC Archaeoglobaceae; Archaeoglobus.
 OX NCBI_TaxID=2234;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
 RX MEDLINE=98049343; PubMed=9389475;
 RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
 RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
 RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,

RA Kirtness E.P., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
 RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
 RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
 RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
 RA Venter J.C.;
 RT "The complete genome sequence of the hyperthermophilic, sulphate-
 RT reducing archaeon *Archaeoglobus fulgidus*.";
 RL Nature 390:364-370(1997).
 DR EMBL; AE001109; AAB91255.1; -;
 DR TIGR; AF2411; -;
 DR InterPro; IPR002103; Bac_luciferase.
 DR Pfam; PF00296; bac_luciferase; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 363 AA; 41736 MW; 0E976EAE788F4803 CRC64;
 Query Match 70.4%; Score 38; DB 17; Length 363;
 Best Local Similarity 54.5%; Pred. No. 16;
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 Qy 1 BEVVPXGMDYS 11
 ||:|:|:|
 Db 120 ENIVPYGIDFS 130
 RESULT 4
 Q8DIHO PRELIMINARY; PRT; 1044 AA.
 ID Q8DIHO;
 AC Q8DIHO;
 DT 01-MAR-2003 (TRENBLrel. 23, Created)
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE Multidrug efflux transporter.
 GN TLL1618.
 OS Synechococcus elongatus (Thermosynechococcus elongatus).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
 OX NCBI_TaxID=32046;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BP-1.
 RX MEDLINE=2225144; PubMed=12240834;
 RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
 RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
 RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
 RA Shimpo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the thermophilic cyanobacterium
 RT *Thermosynechococcus elongatus* BP-1.";
 RL DNA Res. 9:123-130(2002).
 DR EMBL; AP005374; BAC09170.1; -;
 KW Complete proteome.
 SQ SEQUENCE 1044 AA; 113205 MW; 00E9C13F0F636D2F CRC64;
 Query Match 70.4%; Score 38; DB 16; Length 1044;
 Best Local Similarity 63.6%; Pred. No. 53;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 Qy 1 BEVVPXGMDYS 11
 ||:|:|:|
 Db 843 BEVLPNGIGYS 853
 RESULT 5
 O22081 PRELIMINARY; PRT; 341 AA.
 ID O22081;
 AC O22081;
 DT 01-JAN-1998 (TRENBLrel. 05, Created)
 DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
 DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
 DE Sucrose-phosphate synthase (Fragment).
 GN CITS92.
 OS Citrus unshiu (Satsuma orange).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Sapindales; Rutaceae; Citrus.
 OX NCBI_TaxID=55188;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Miyagawa-Wase; TISSUE=Juice sacs and segment epidermis;
 RX MEDLINE=96439842; PubMed=8842155;
 RA Komatsu A., Takanokura Y., Omura M., Akihama T.;
 RT "Cloning and molecular analysis of cDNA encoding three sucrose
 RT phosphate synthase isoforms from a citrus fruit (Citrus unshiu
 RT Marc.).";
 RL Mol. Gen. Genet. 252:346-351 (1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Miyagawa-Wase; TISSUE=Juice sacs and segment epidermis;
 RA Komatsu A., Takanokura Y., Moriguchi T., Omura M., Akihama T.;
 RT "Differential expression of three sucrose-phosphate synthase isoforms
 RT during sucrose accumulation in citrus fruits (Citrus unshiu Marc.).";
 RL Plant Sci. 140:169-178 (1999).
 DR EMBL; AB006319; BAA23215.1; -.
 FT NON_TER 1 341
 FT NON_TER 341 341
 SQ SEQUENCE 341 AA; 36136 MW; 61417A69C4560777 CRC64;
 Query Match 66.7%; Score 36; DB 10; Length 341;
 Best Local Similarity 66.7%; Pred. No. 40;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 3 VVPXGMDYS 11
 Db [1] [1] [1] [1]
 Db 228 VPPGMDFS 236
 RESULT 6
 ID O22096 PRELIMINARY; PRT; 348 AA.
 AC O22096;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Sucrose-phosphate synthase (Fragment).
 GN CITSPS3.
 OS Citrus unshiu (Satsuma orange).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Sapindales; Rutaceae; Citrus.
 OX NCBI_TaxID=55188;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Juice sacs and segment epidermis;
 RX MEDLINE=96439842; PubMed=8842155;
 RA Komatsu A., Takanokura Y., Omura M., Akihama T.;
 RT "Cloning and molecular analysis of cDNA encoding three sucrose
 RT phosphate synthase isoforms from a citrus fruit (Citrus unshiu
 RT Marc.).";
 RL Mol. Gen. Genet. 252:346-351 (1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Juice sacs and segment epidermis;
 RA Komatsu A., Takanokura Y., Moriguchi T., Omura M., Akihama T.;
 RT "Differential expression of three sucrose-phosphate synthase isoforms
 RT during sucrose accumulation in citrus fruits (Citrus unshiu Marc.).";
 RL Plant Sci. 140:169-178 (1999).
 DR EMBL; AB006660; BAA22071.1; -.
 FT NON_TER 1 348
 FT NON_TER 348 348
 SQ SEQUENCE 348 AA; 36556 MW; BE1C21BA6FF5C5E CRC64;
 Query Match 66.7%; Score 36; DB 10; Length 348;
 Best Local Similarity 66.7%; Pred. No. 41;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 3 VVPXGMDYS 11

Db 234 VPPGMDFS 242
 [1] [1] [1] [1]
 ID O22096 PRELIMINARY; PRT; 452 AA.
 AC O22096;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Atg7750/F25P22.17.
 GN Atg7750/F25P22.17.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Chen H., Kim C.J., Koesema B., Meyers M.C., Banh J.,
 RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
 RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
 RA Kawai J., Lam J., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
 RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
 RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
 RA Ecker J.R.;
 RT "Arabidopsis cDNA clones.";
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Kim C.J., Chen H., Cheuk R., Shinn P., Banh J., Bowser L.,
 RA Carninci P., Chang E., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
 RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
 RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,
 RA Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
 RA Tang C.C., Toriumi M., Wu H.C., Yamada K., Yamamura Y., Yu G., Yu S.,
 RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
 RT "Arabidopsis ORF clones.";
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF419606; AAL31938.1; -.
 DR EMBL; AY113044; AAM47352.1; -.
 DR InterPro; IPR002471; Prol endopep. ser.
 DR InterPro; IPR000379; Ser esterase.
 DR PROSITE; PS00708; PRO ENDOPEP_SER.1
 SQ SEQUENCE 452 AA; 49682 MW; A159955B21742C4A CRC64;
 Query Match 66.7%; Score 36; DB 10; Length 452;
 Best Local Similarity 70.0%; Pred. No. 55;
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 BEVFXGMDY 10
 Db [1] [1] [1] [1]
 Db 210 BEDVFSAMDY 219
 RESULT 8
 ID O9C9T7 PRELIMINARY; PRT; 460 AA.
 AC O9C9T7;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical 50.6 kDa protein.
 GN F25P22.17.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;


```

RESULT 10
Q9SN30
ID Q9SN30 PRELIMINARY; PRT; 1083 AA.
AC Q9SN30;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Sucrose-phosphate synthase-like protein (EC 2.4.1.14).
GN F28M11.40 OR AT4G10120.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OC NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Murphy G., Ridley P., Hudson S., Bancroft I., Mewes H.W.,
RA Mayer K.F.X., Lemcke K., Schueller C.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL049487; CAB39764.1; -
DR EMBL; AL161516; CAB78135.1; -
DR InterPro; IPR001296; Glyco_trans_1.
DR Pfam; PF00534; Glycosyl_transf_1; 1.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 1083 AA; 122688 MW; EFCl26FCA2137BB1 CRC64;

Query Match 66.7%; Score 36; DB 10; Length 1083;
Best Local Similarity 66.7%; Pred. NO. 1.5e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
|:| |||:|
DB 483 VVPGMDFS 491

RESULT 11
Q43010
ID Q43010 PRELIMINARY; PRT; 1084 AA.
AC Q43010;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Sucrose phosphate synthase.
GN SPS1.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoidae; Oryzaceae; Oryza.
OC NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. japonica;
RA Sakamoto M., Satozawa T., Kishimoto N., Higo K., Shimada H.,
RA Fujimura T.;
RL "Structure and RFLP mapping of a rice sucrose phosphate synthase (SPS) gene that is specifically expressed in the source organ.";
RT Plant Sci. 112:207-217(1995).
DR EMBL; D45890; BAA08304.1; -
DR Gramene; Q43010; -
DR InterPro; IPR001296; Glyco_trans_1.
DR InterPro; IPR001557; L_LDH.
DR Pfam; PF00534; Glycosyl_transf_1; 1.
DR PROSITE; PS00064; L_LDH; 1.
SQ SEQUENCE 1084 AA; 119417 MW; 4D0D79AFEDD2F92B CRC64;

```

Query Match 66.7%; Score 36; DB 10; Length 1084;
 Best Local Similarity 66.7%; Pred. No. 1.5e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
 |:|:|:|:
 Db 453 VIPPGMDFS 461

RESULT 12

Q8S064 Q8S064 PRELIMINARY; PRT; 1100 AA.
 AC Q8S064;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-WAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Putative sucrose-phosphate synthase.
 GN P0678F11.14.
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzeae; Oryza.
 OX NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
 RT clone:P0678F11.1";
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP003437; BAB86107.1; -;
 DR Gramene; Q8S064; -;
 DR InterPro; IPR001296; Glyco_trans_1.
 DR InterPro; IPR001557; L_LDH.
 DR Pfam; PF00534; Glycos_transf_1; 1.
 DR PROSITE; PS00064; L_LDH; 1.
 SQ SEQUENCE 1100 AA; 121170 MW; 084P4604BA389CAD CRC64;

Query Match 66.7%; Score 36; DB 10; Length 1100;
 Best Local Similarity 66.7%; Pred. No. 1.5e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 VVPXGMDYS 11
 |:|:|:|:
 Db 459 VIPPGMDFS 477

RESULT 13

Q9GQ04 Q9GQ04 PRELIMINARY; PRT; 219 AA.
 AC Q9GQ04;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
 DE Na+/K+/2Cl-cotransporter (Fragment).
 OS Eriocheir sinensis (Chinese mitten crab).
 OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
 OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
 OC Eubrachyura; Grapsoidae; Varunidae; Eriocheir.
 OX NCBI_TaxID=95602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Gill;
 RA Wehrhau D., Towle D.W.;
 RT "Na+/H+-exchanger and Na+/K+/2Cl- cotransporter are expressed in
 RT gills of the euryhaline Chinese crab Eriocheir sinensis";
 RL Comp. Biochem. Physiol. 126:S158-S158(2000).
 DR EMBL; AF301160; AAG39938.1; -;
 FT NON TER 1
 FT NON TER 219
 SQ SEQUENCE 219 AA; 24159 MW; 599442DA36F0D3F1 CRC64;

Query Match 64.8%; Score 35; DB 5; Length 219;
 Best Local Similarity 75.0%; Pred. No. 39;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 VVPXGMDYS 11
 |:|:|:|:
 Db 107 VPOGLDYS 114

RESULT 14

Q8XPAB Q8XPAB PRELIMINARY; PRT; 253 AA.
 AC Q8XPAB;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Hypothetical protein CPE0057.
 GN CPE0057.
 OS Clostridium perfringens.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1502;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=13 / Type A;
 RX PubMed=11792842;
 RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
 RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
 RT "Complete genome sequence of Clostridium perfringens, an anaerobic
 RT flesh-eater.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
 DR EMBL; AP003185; BAB79763.1; -;
 DR InterPro; IPR000205; NAD_binding.
 DR InterPro; IPR000594; Thif_domain.
 DR Pfam; PF00899; Thif; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 253 AA; 27714 MW; 814DF9D3E0D7486 CRC64;

Query Match 64.8%; Score 35; DB 16; Length 253;
 Best Local Similarity 50.0%; Pred. No. 46;
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGMDY 10
 |:|:|:|:
 Db 108 BEIIPDDVDY 117

RESULT 15

O52367 O52367 PRELIMINARY; PRT; 298 AA.
 AC O52367;
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DB Aryl-alcohol dehydrogenase homolog (Fragment).
 GN XVLB1
 OS Rhizobium tropici.
 OG Plasmid pRtrCFN299A.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Rhizobium.
 OX NCBI_TaxID=398;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CFN299;
 RA Rosenbluth M., Hynes M.F., Martinez-Romero E.;
 RT Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- COPACTOR: ZINC (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE
 CC FAMILY.
 DR EMBL; AF036920; AAC04779.1; -;
 DR InterPro; IPR002328; ADH_zinc.
 DR InterPro; IPR002085; Adh_zn_family.
 DR InterPro; IPR000205; NAD_binding.

DR Pfam; PF00107; adh zinc; 1.
DR PROSITE; PS00059; ADH ZINC; 1.
KW Metal-binding; Oxidoreductase; Zinc; Plasmid.
FT NON_TER 298
SQ SEQUENCE 298 AA; 31092 MW; 49E2F8117C33AE87 CRC64;
Query Match 64.8%; Score 35; DB 2; Length 298;
Best Local Similarity 50.0%; Pred. No. 55;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
Oy 2 EVVPXGMDYS 11
Db 250 EIIPGADFS 259

Search completed: December 22, 2003, 17:51:30
Job time : 26.2667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:41:00 ; Search time 32.4667 Seconds
(without alignments)
53.778 Million cell updates/sec

Title: US-09-909-164-42

Perfect score: 52

Sequence: 1 EVVVRGSHYS(11)

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq_19Jun03.*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|----------------------------|
| 1 | 49 | 94.2 | 11 | 23 | ABB80546 Hepatitis C virus |
| 2 | 49 | 94.2 | 11 | 23 | ABB80550 Hepatitis C virus |
| 3 | 49 | 94.2 | 11 | 23 | ABB80554 Hepatitis C virus |
| 4 | 49 | 94.2 | 11 | 23 | ABB80555 Hepatitis C virus |
| 5 | 48 | 92.3 | 11 | 23 | ABB80523 Hepatitis C virus |
| 6 | 48 | 92.3 | 11 | 23 | ABB80527 Hepatitis C virus |
| 7 | 48 | 92.3 | 11 | 23 | ABB80531 Hepatitis C virus |
| 8 | 48 | 92.3 | 11 | 23 | ABB80532 Hepatitis C virus |
| 9 | 48 | 92.3 | 11 | 23 | ABB80537 Hepatitis C virus |

| | | | | | |
|----|----|------|----|----|----------------------------|
| 10 | 48 | 92.3 | 11 | 23 | ABB80541 Hepatitis C virus |
| 11 | 48 | 92.3 | 11 | 23 | ABB80558 Hepatitis C virus |
| 12 | 48 | 92.3 | 11 | 23 | ABB80560 Hepatitis C virus |
| 13 | 40 | 76.9 | 11 | 23 | ABB80544 Hepatitis C virus |
| 14 | 40 | 76.9 | 11 | 23 | ABB80545 Hepatitis C virus |
| 15 | 40 | 76.9 | 11 | 23 | ABB80547 Hepatitis C virus |
| 16 | 40 | 76.9 | 11 | 23 | ABB80548 Hepatitis C virus |
| 17 | 40 | 76.9 | 11 | 23 | ABB80549 Hepatitis C virus |
| 18 | 40 | 76.9 | 11 | 23 | ABB80551 Hepatitis C virus |
| 19 | 40 | 76.9 | 11 | 23 | ABB80552 Hepatitis C virus |
| 20 | 40 | 76.9 | 11 | 23 | ABB80553 Hepatitis C virus |
| 21 | 40 | 76.9 | 11 | 23 | ABB80556 Hepatitis C virus |
| 22 | 40 | 76.9 | 11 | 23 | ABB80557 Hepatitis C virus |
| 23 | 39 | 75.0 | 11 | 23 | ABB80521 Hepatitis C virus |
| 24 | 39 | 75.0 | 11 | 23 | ABB80522 Hepatitis C virus |
| 25 | 39 | 75.0 | 11 | 23 | ABB80524 Hepatitis C virus |
| 26 | 39 | 75.0 | 11 | 23 | ABB80525 Hepatitis C virus |
| 27 | 39 | 75.0 | 11 | 23 | ABB80526 Hepatitis C virus |
| 28 | 39 | 75.0 | 11 | 23 | ABB80528 Hepatitis C virus |
| 29 | 39 | 75.0 | 11 | 23 | ABB80529 Hepatitis C virus |
| 30 | 39 | 75.0 | 11 | 23 | ABB80530 Hepatitis C virus |
| 31 | 39 | 75.0 | 11 | 23 | ABB80533 Hepatitis C virus |
| 32 | 39 | 75.0 | 11 | 23 | ABB80534 Hepatitis C virus |
| 33 | 39 | 75.0 | 11 | 23 | ABB80535 Hepatitis C virus |
| 34 | 39 | 75.0 | 11 | 23 | ABB80536 Hepatitis C virus |
| 35 | 39 | 75.0 | 11 | 23 | ABB80538 Hepatitis C virus |
| 36 | 39 | 75.0 | 11 | 23 | ABB80539 Hepatitis C virus |
| 37 | 39 | 75.0 | 11 | 23 | ABB80540 Hepatitis C virus |
| 38 | 39 | 75.0 | 11 | 23 | ABB80542 Hepatitis C virus |
| 39 | 39 | 75.0 | 11 | 23 | ABB80543 Hepatitis C virus |
| 40 | 39 | 75.0 | 11 | 23 | ABB80559 Hepatitis C virus |
| 41 | 39 | 75.0 | 11 | 23 | ABB80561 Hepatitis C virus |
| 42 | 39 | 75.0 | 11 | 23 | ABB80562 Hepatitis C virus |
| 43 | 39 | 75.0 | 11 | 23 | ABB80563 Hepatitis C virus |
| 44 | 39 | 75.0 | 11 | 23 | ABB80564 Hepatitis C virus |
| 45 | 39 | 75.0 | 11 | 23 | ABB80565 Hepatitis C virus |

ALIGNMENTS

RESULT 1

ABB80546

ID ABB80546 standard; peptide; 11 AA.

AC ABB80546;

DT 08-OCT-2002 (first entry)

DB Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #26.

KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide; virucide.

OS Synthetic.

PH Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"

FT Modified-site 11 /note= "C-terminal amide"

XX WO200208251-A2.

XX 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US23169.

XX 21-JUL-2000; 2000US-220101P.

XX

CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.

SQ Sequence 11 AA;

Query Match 94.2%; Score 49; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0037;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXGSHYS 11
| | | | | | | |
DB 1 BEVVPXGSHYS 11

RESULT 4
ABB80555
ID ABB80555 standard; peptide; 11 AA.
XX
AC ABB80555;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #35.
XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX
OS Synthetic.

PH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Misc-difference 9 /note= "D-form residue"
FT Modified-site 11 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"

PN WO200208251-A2.

XX 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US23169.

XX 21-JUL-2000; 2000US-220101P.

XX (CORV-) CORVAS INT INC.

XX Lim-wilby M, Levy OE, Brunck TK;

XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -

PS Claim 17; Page 65; 69pp; English.

XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.

XX Sequence 11 AA;

Query Match 94.2%; Score 49; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0037;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXGSHYS 11
| | | | | | | |
DB 1 BEVVPXGSHYS 11

RESULT 5
ABB80523
ID ABB80523 standard; peptide; 11 AA.

XX AC ABB80523;

XX 08-OCT-2002 (first entry)

XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #3.

XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.

XX Synthetic.

PH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"

FT Misc-difference 9 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

XX WO200208251-A2.

XX 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US23169.

XX 21-JUL-2000; 2000US-220101P.

XX (CORV-) CORVAS INT INC.

XX Lim-wilby M, Levy OE, Brunck TK;

XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -

PS Claim 17; Page 64; 69pp; English.

XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.

XX Sequence 11 AA;

Query Match 92.3%; Score 48; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0058;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXGSHYS 11
| | | | | | | |
DB 1 BEVVPXGSHYS 11

```

RESULT 6
ABB80527
ID ABB80527 standard; peptide; 11 AA.
XX AC ABB80527;
XX DT 08-OCT-2002 (first entry)
XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #7.
XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX OS virucide.
XX OS Synthetic.
XX FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX WO200208251-A2.
XX 31-JAN-2002.
XX 19-JUL-2001; 2001WO-US23169.
XX 21-JUL-2000; 2000US-220101P.
XX (CORV-) CORVAS INT INC.
XX Lim-wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX Novel peptide compound having hepatitis C virus protease inhibitory
FT activity useful for treating disorders associated with hepatitis C
FT virus protease -
XX Claim 17; Page 64; 69pp; English.
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX Sequence 11 AA;
XX Query Match 92.3%; Score 48; DB 23; Length 11;
XX Best Local Similarity 90.9%; Pred. NO. 0.0058;
XX Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGXHYS 11
DB 1 EEVVPXGXHYS 11
RESULT 7
ABB80531
ID ABB80531 standard; peptide; 11 AA.
XX AC ABB80531;
XX DT 08-OCT-2002 (first entry)
XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #12.
XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX OS virucide.
XX OS Synthetic.
XX FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX WO200208251-A2.
XX 31-JAN-2002.
XX 19-JUL-2001; 2001WO-US23169.
XX 21-JUL-2000; 2000US-220101P.
XX (CORV-) CORVAS INT INC.
XX Lim-wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX Novel peptide compound having hepatitis C virus protease inhibitory
FT activity useful for treating disorders associated with hepatitis C
FT virus protease -
XX Claim 17; Page 64; 69pp; English.
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX Sequence 11 AA;
XX Query Match 92.3%; Score 48; DB 23; Length 11;
XX Best Local Similarity 90.9%; Pred. NO. 0.0058;
XX Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGXHYS 11
DB 1 EEVVPXGXHYS 11

```

```

XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #11.
XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX OS Synthetic.
XX FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Modified-site 11 /note= "C-terminal amide"
XX WO200208251-A2.
XX 31-JAN-2002.
XX 19-JUL-2001; 2001WO-US23169.
XX 21-JUL-2000; 2000US-220101P.
XX (CORV-) CORVAS INT INC.
XX Lim-wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX Novel peptide compound having hepatitis C virus protease inhibitory
FT activity useful for treating disorders associated with hepatitis C
FT virus protease -
XX Claim 17; Page 64; 69pp; English.
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX Sequence 11 AA;
XX Query Match 92.3%; Score 48; DB 23; Length 11;
XX Best Local Similarity 90.9%; Pred. NO. 0.0058;
XX Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGXHYS 11
DB 1 EEVVPXGXHYS 11
RESULT 8
ABB80532
ID ABB80532 standard; peptide; 11 AA.
XX AC ABB80532;
XX DT 08-OCT-2002 (first entry)
XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #12.
XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX OS virucide.
XX OS Synthetic.
XX FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX WO200208251-A2.
XX 31-JAN-2002.
XX 19-JUL-2001; 2001WO-US23169.
XX 21-JUL-2000; 2000US-220101P.
XX (CORV-) CORVAS INT INC.
XX Lim-wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX Novel peptide compound having hepatitis C virus protease inhibitory
FT activity useful for treating disorders associated with hepatitis C
FT virus protease -
XX Claim 17; Page 64; 69pp; English.
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX Sequence 11 AA;
XX Query Match 92.3%; Score 48; DB 23; Length 11;
XX Best Local Similarity 90.9%; Pred. NO. 0.0058;
XX Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGXHYS 11
DB 1 EEVVPXGXHYS 11
RESULT 8
ABB80532
ID ABB80532 standard; peptide; 11 AA.
XX AC ABB80532;
XX DT 08-OCT-2002 (first entry)
XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #12.
XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX OS virucide.
XX OS Synthetic.
XX FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX WO200208251-A2.
XX 31-JAN-2002.
XX 19-JUL-2001; 2001WO-US23169.
XX 21-JUL-2000; 2000US-220101P.
XX (CORV-) CORVAS INT INC.
XX Lim-wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX Novel peptide compound having hepatitis C virus protease inhibitory
FT activity useful for treating disorders associated with hepatitis C
FT virus protease -
XX Claim 17; Page 64; 69pp; English.
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX Sequence 11 AA;
XX Query Match 92.3%; Score 48; DB 23; Length 11;
XX Best Local Similarity 90.9%; Pred. NO. 0.0058;
XX Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 EEVVPXGXHYS 11
DB 1 EEVVPXGXHYS 11

```

```

FT Modified-site 6 /note= "N-terminal acetyl"
FT FT /note= "Norvalyl carbonyl forming keto-amide linkage with
FT FT residue 7"
FT
FT Misc-difference 9 /note= "D-form residue"
FT FT 11
FT Modified-site 11 /note= "C-terminal amide"
FT FT
FT WO200208251-A2.
FT
FT 31-JAN-2002.
FT
FT 19-JUL-2001; 2001WO-US23169.
FT
FT 21-JUL-2000; 2000US-220101P.
FT
FT (CORV-) CORVAS INT INC.
FT
FT Lim-wilby M, Levy OE, Brunck TK;
FT
FT WPI; 2002-361643/39.
FT
FT Novel peptide compound having hepatitis C virus protease inhibitory
FT activity useful for treating disorders associated with hepatitis C
FT virus protease
FT
FT Claim 17; Page 64; 69pp; English.
FT
FT The sequence represents a peptide compound of the invention having
FT hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
FT invention are alpha-ketoamide peptide analogues. The peptides have
FT virucide activity, and are useful for treating and in the manufacture of
FT a medicament to treat disorders associated with HCV protease. A
FT pharmaceutical composition comprising the peptide as an active ingredient
FT is useful for treating disorders associated with hepatitis C virus.
FT
FT SQ Sequence 11 AA;
FT
FT Query Match 92.3%; Score 48; DB 23; Length 11;
FT Best Local Similarity 90.9%; Pred. No. 0.0058;
FT Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
FT
FT QY 1 EEVVPXGXHY 11
FT |||||
FT Db 1 EEVVPXGXHY 11
FT
FT RESULT 9
FT ABB80537
FT ID ABB80537 standard; peptide; 11 AA.
FT
FT AC ABB80537;
FT
FT DT 08-OCT-2002 (first entry)
FT
FT DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #17.
FT
FT KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
FT virucide.
FT
FT OS Synthetic.
FT
FT Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT FT 6
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT FT residue 7"
FT
FT Misc-difference 9 /note= "D-form residue"
FT FT 11
FT Modified-site 11 /note= "C-terminal amide"
FT

```

```

XX WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US23169.
XX
XX 21-JUL-2000; 2000US-220101P.
XX
XX (CORV-) CORVAS INT INC.
XX
XX Lim-wilby M, Levy OE, Brunck TK;
XX
XX WPI; 2002-361643/39.
XX
XX Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C
XX virus protease
XX
XX Claim 17; Page 64; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX invention are alpha-ketoamide peptide analogues. The peptides have
XX virucide activity, and are useful for treating and in the manufacture of
XX a medicament to treat disorders associated with HCV protease. A
XX pharmaceutical composition comprising the peptide as an active ingredient
XX is useful for treating disorders associated with hepatitis C virus.
XX
XX SQ Sequence 11 AA;
XX
XX Query Match 92.3%; Score 48; DB 23; Length 11;
XX Best Local Similarity 90.9%; Pred. No. 0.0058;
XX Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 EEVVPXGXHY 11
XX |||||
XX Db 1 EEVVPXGXHY 11
XX
XX RESULT 10
XX ABB80541
XX ID ABB80541 standard; peptide; 11 AA.
XX
XX AC ABB80541;
XX
XX DT 08-OCT-2002 (first entry)
XX
XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #21.
XX
XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX
XX OS Synthetic.
XX
XX Key Location/Qualifiers
XX Modified-site 1 /note= "N-terminal acetyl"
XX FT 6
XX Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
XX FT residue 7"
XX
XX Misc-difference 8 /note= "D-form residue"
XX FT 11
XX Modified-site 11 /note= "C-terminal amide"
XX
XX WO200208251-A2.
XX
XX 31-JAN-2002.
XX
XX 19-JUL-2001; 2001WO-US23169.
XX
XX 21-JUL-2000; 2000US-220101P.
XX

```


XX PA (CORV-) CORVAS INT INC.
 XX PI Lim-wilby M, Levy OE, Brunck TK;
 XX DR WPI; 2002-361643/39.
 XX CC The sequence represents a peptide compound of the invention having
 PT Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C
 PT virus protease -
 XX PS Claim 17; Page 65; 69pp; English.
 XX CC The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.
 XX SQ Sequence 11 AA;
 Query Match 92.3%; Score 48; DB 23; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.0058;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 BEVVPXGXHYS 11
 |||||
 Db 1 BEVVPXGXHYS 11
 |||||
 RESULT 11
 ABB80558
 ID ABB80558 standard; peptide; 11 AA.
 XX AC ABB80558;
 XX DT 08-OCT-2002 (first entry)
 XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #38.
 XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 XX KW virucide.
 XX OS Synthetic.
 XX FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Modified-site 8 /note= "Oxymethionine"
 FT Modified-site 11 /note= "C-terminal amide"
 XX WO200208251-A2.
 XX 31-JAN-2002.
 XX 19-JUL-2001; 2001WO-US23169.
 XX 21-JUL-2000; 2000US-220101P.
 XX (CORV-) CORVAS INT INC.
 XX Lim-wilby M, Levy OE, Brunck TK;
 XX WPI; 2002-361643/39.
 XX Novel peptide compound having hepatitis C virus protease inhibitory

PT activity useful for treating disorders associated with hepatitis C
 PT virus protease -
 XX PS Claim 17; Page 65; 69pp; English.
 XX CC The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.
 XX SQ Sequence 11 AA;
 Query Match 92.3%; Score 48; DB 23; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.0058;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 BEVVPXGXHYS 11
 |||||
 Db 1 BEVVPXGXHYS 11
 |||||
 RESULT 12
 ABB80560
 ID ABB80560 standard; peptide; 11 AA.
 XX AC ABB80560;
 XX DT 08-OCT-2002 (first entry)
 XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #40.
 XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 XX KW virucide.
 XX OS Synthetic.
 XX FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Misc-difference 8 /note= "D-form residue"
 FT Modified-site 8 /note= "Oxymethionine"
 FT Misc-difference 9 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"
 XX WO200208251-A2.
 XX 31-JAN-2002.
 XX 19-JUL-2001; 2001WO-US23169.
 XX 21-JUL-2000; 2000US-220101P.
 XX (CORV-) CORVAS INT INC.
 XX Lim-wilby M, Levy OE, Brunck TK;
 XX WPI; 2002-361643/39.
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C
 PT virus protease -
 XX PS Claim 17; Page 65; 69pp; English.

XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.

SQ Sequence 11 AA;

Query Match 92.3%; Score 48; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.0058;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 BEVVPXGXHYS 11
| | | | | | | | | |
Db 1 BEVVPXGMHYS 11

RESULT 13

ABB80544
ID ABB80544 standard; peptide; 11 AA.

XX AC ABB80544;

XX DT 08-OCT-2002 (first entry)

XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #24.

XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.

XX OS Synthetic.

XX FH Key Location/Qualifiers

XX FT Modified-site 1

XX FT /note= "N-terminal acetyl"

XX FT Modified-site 6

XX FT /note= "Norvalyl carbonyl forming keto-amide linkage with
XX residue 7"

XX FT Modified-site 11

XX FT /note= "C-terminal amide"

XX WO200208251-A2.

XX PD 31-JAN-2002.

XX PP 19-JUL-2001; 2001WO-US23169.

XX PR 21-JUL-2000; 2000US-220101P.

XX PA (CORV-) CORVAS INT INC.

XX PI Lin-wilby M, Levy OE, Brunck TK;

XX DR WPI; 2002-361643/39.

XX FT Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C
XX virus protease

XX PS Claim 17; Page 65; 69pp; English.

XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.

SQ Sequence 11 AA;

Query Match 76.9%; Score 40; DB 23; Length 11;
Best Local Similarity 81.8%; Pred. No. 0.23;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 BEVVPXGXHYS 11
| | | | | | | | | |
Db 1 BEVVPXGTSYS 11

RESULT 14

ABB80545
ID ABB80545 standard; peptide; 11 AA.

XX AC ABB80545;

XX DT 08-OCT-2002 (first entry)

XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #25.

XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.

XX OS Synthetic.

XX FH Key Location/Qualifiers

XX FT Modified-site 1

XX FT /note= "N-terminal acetyl"

XX FT Modified-site 6

XX FT /note= "Norvalyl carbonyl forming keto-amide linkage with
XX residue 7"

XX FT Misc-difference 9

XX FT /note= "D-form residue"

XX FT Modified-site 11

XX FT /note= "C-terminal amide"

XX WO200208251-A2.

XX PD 31-JAN-2002.

XX PP 19-JUL-2001; 2001WO-US23169.

XX PR 21-JUL-2000; 2000US-220101P.

XX PA (CORV-) CORVAS INT INC.

XX PI Lin-wilby M, Levy OE, Brunck TK;

XX DR WPI; 2002-361643/39.

XX FT Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C
XX virus protease

XX PS Claim 17; Page 65; 69pp; English.

XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.

SQ Sequence 11 AA;

Query Match 76.9%; Score 40; DB 23; Length 11;
Best Local Similarity 81.8%; Pred. No. 0.23;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 BEVVPXGXHYS 11
| | | | | | | | | |
Db 1 BEVVPXGTSYS 11

```

RESULT 15
ABB80547
ID ABB80547 standard; peptide; 11 AA.
XX AC ABB80547;
XX DT 08-OCT-2002 (first entry)
XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #27.
XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX OS virucide.
XX OS Synthetic.
XX FT Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Modified-site 11 /note= "C-terminal amide"
XX WO200208251-A2.
XX 31-JAN-2002.
XX 19-JUL-2001; 2001WO-US23169.
XX 21-JUL-2000; 2000US-220101P.
XX (CORV-) CORVAS INT INC.
XX Lim-wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX Novel peptide compound having hepatitis C virus protease inhibitory
XX activity useful for treating disorders associated with hepatitis C
XX virus protease -
XX Claim 17; Page 65; 69pp; English.
XX The sequence represents a peptide compound of the invention having
XX hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
XX invention are alpha-ketoamide peptide analogues. The peptides have
XX virucide activity, and are useful for treating and in the manufacture of
XX a medicament to treat disorders associated with HCV protease. A
XX pharmaceutical composition comprising the peptide as an active ingredient
XX is useful for treating disorders associated with hepatitis C virus.
SQ Sequence 11 AA;
Query Match 76.9%; Score 40; DB 23; Length 11;
Best Local Similarity 81.8%; Pred. No. 0.23;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 EEVVPXGXHYS 11
Db 1 EEVVPXGTDYS 11
Search completed: December 22, 2003, 17:41:02
Job time : 33.4667 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:37:03 ; Search time 10.1333 Seconds
(without alignments)
45.930 Million cell updates/sec

Title: US-09-909-164-42
Perfect score: 52
Sequence: 1 REVVPXGXHY 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgm2_6/ptodata/1/iaa/5A COMB.pep:*
2: /cgm2_6/ptodata/1/iaa/5B COMB.pep:*
3: /cgm2_6/ptodata/1/iaa/6A COMB.pep:*
4: /cgm2_6/ptodata/1/iaa/6B COMB.pep:*
5: /cgm2_6/ptodata/1/iaa/PTUS COMB.pep:*
6: /cgm2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|----------------------|
| 1 | 38 | 73.1 | 856 | 4 | US-09-252-991A-21444 |
| 2 | 36 | 69.2 | 1037 | 4 | US-09-134-001C-4794 |
| 3 | 35 | 67.3 | 277 | 4 | US-09-252-991A-26615 |
| 4 | 34 | 65.4 | 126 | 2 | US-08-879-995A-3 |
| 5 | 34 | 65.4 | 126 | 3 | US-09-215-096-3 |
| 6 | 33 | 63.5 | 63 | 4 | US-09-107-532A-6771 |
| 7 | 33 | 63.5 | 447 | 3 | US-08-961-083-182 |
| 8 | 33 | 63.5 | 447 | 4 | US-09-536-784-182 |
| 9 | 33 | 63.5 | 484 | 4 | US-09-468-656A-6 |
| 10 | 33 | 63.5 | 600 | 2 | US-08-821-119-19 |
| 11 | 33 | 63.5 | 600 | 2 | US-08-821-118-2 |
| 12 | 33 | 63.5 | 763 | 3 | US-08-961-083-66 |
| 13 | 33 | 63.5 | 763 | 4 | US-09-536-784-66 |
| 14 | 33 | 63.5 | 796 | 3 | US-08-961-083-56 |
| 15 | 33 | 63.5 | 796 | 4 | US-09-536-784-56 |
| 16 | 33 | 63.5 | 819 | 4 | US-09-468-656A-8 |
| 17 | 33 | 63.5 | 819 | 4 | US-09-468-656A-10 |
| 18 | 33 | 63.5 | 838 | 4 | US-09-468-656A-4 |
| 19 | 32 | 61.5 | 253 | 4 | US-09-328-352-6440 |
| 20 | 32 | 61.5 | 738 | 1 | US-08-530-010-3 |
| 21 | 32 | 61.5 | 738 | 1 | US-08-530-010-5 |
| 22 | 32 | 61.5 | 738 | 1 | US-08-530-010-7 |
| 23 | 32 | 61.5 | 738 | 1 | US-08-530-010-9 |
| 24 | 32 | 61.5 | 738 | 1 | US-08-530-010-11 |
| 25 | 32 | 61.5 | 738 | 2 | US-08-484-101B-3 |
| 26 | 32 | 61.5 | 738 | 2 | US-08-484-101B-5 |
| 27 | 32 | 61.5 | 738 | 2 | US-08-484-101B-7 |

| | | | | | |
|----|------|------|---|----------------------|--------------------|
| 28 | 61.5 | 738 | 2 | US-08-484-101B-9 | Sequence 9, Appli |
| 29 | 61.5 | 738 | 2 | US-08-484-101B-11 | Sequence 11, Appli |
| 30 | 61.5 | 738 | 3 | US-08-714-524D-3 | Sequence 3, Appli |
| 31 | 61.5 | 738 | 3 | US-08-714-524D-5 | Sequence 5, Appli |
| 32 | 61.5 | 738 | 3 | US-08-714-524D-7 | Sequence 7, Appli |
| 33 | 61.5 | 738 | 3 | US-08-714-524D-11 | Sequence 11, Appli |
| 34 | 61.5 | 1247 | 4 | US-09-252-991A-32960 | Sequence 32960, A |
| 35 | 61.5 | 1247 | 4 | US-09-134-001C-3950 | Sequence 3950, Ap |
| 36 | 59.6 | 70 | 4 | US-09-134-001C-4185 | Sequence 4185, Ap |
| 37 | 59.6 | 237 | 4 | US-09-328-352-7284 | Sequence 7284, Ap |
| 38 | 59.6 | 252 | 4 | US-08-178-257-15 | Sequence 15, Appl |
| 39 | 59.6 | 388 | 4 | US-09-342-647-4 | Sequence 4, Appli |
| 40 | 59.6 | 502 | 2 | US-08-459-146-2 | Sequence 2, Appli |
| 41 | 59.6 | 622 | 2 | US-08-459-065-2 | Sequence 2, Appli |
| 42 | 59.6 | 622 | 2 | US-08-459-065-2 | Sequence 2, Appli |
| 43 | 59.6 | 630 | 4 | US-09-342-647-2 | Sequence 2, Appli |
| 44 | 59.6 | 667 | 4 | US-09-342-647-28 | Sequence 28, Appli |
| 45 | 59.6 | 700 | 4 | US-09-621-816B-2 | Sequence 2, Appli |

ALIGNMENTS

RESULT 1:
US-09-252-991A-21444
; Sequence 21444, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21444
; LENGTH: 856
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21444

Query Match 73.1%; Score 38; DB 4; Length 856;
Best Local Similarity 70.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 REVVPXGXHY 10
Db 64 EAVVPGGEHY 73

RESULT 2

US-09-134-001C-4794
; Sequence 4794, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCU
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4794
; LENGTH: 1037
; TYPE: PRT

ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4794

Query Match 69.2%; Score 36; DB 4; Length 1037;
Best Local Similarity 63.6%; Pred. No. 47;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXHYS 11
:|||||
Db 199 KEVVSNGLHYS 209

RESULT 3
US-09-252-991A-26615
Sequence 26615, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252.991A
CURRENT FILING DATE: 1999-02-18
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 26615
LENGTH: 277
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26615

Query Match 67.3%; Score 35; DB 4; Length 277;
Best Local Similarity 63.6%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 BEVVPXGXHYS 11
:|||||
Db 48 EETVPGGHTS 58

RESULT 4
US-08-879-995A-3
Sequence 3, Application US/08879995A
Patent No. 5985606
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Kaser, Matthew R.
TITLE OF INVENTION: HUMAN PREPROTACHYKININ B
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/879.995A
FILING DATE: Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
PRIOR APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0326 US
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 163590
US-08-879-995A-3

Query Match 65.4%; Score 34; DB 2; Length 126;
Best Local Similarity 66.7%; Pred. No. 12;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGXH 9
:|||||
Db 28 EQVVPGGGH 36

RESULT 5
US-09-215-096-3
Sequence 3, Application US/09215096
Patent No. 6008194
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Kaser, Matthew R.
TITLE OF INVENTION: HUMAN PREPROTACHYKININ B
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/215.096
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/879.995
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0326 US
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 163590

US-09-215-096-3

Query Match 65.4%; Score 34; DB 3; Length 126;
 Best Local Similarity 66.7%; Pred. No. 12;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 BEVVPXGXH 9
 Db 28 EQVVEGGGH 36

RESULT 6

US-09-107-532A-6771
 ; Sequence 6771, Application US/09107532A
 ; Patent No. 6583275
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn A Doucette-Stamm and David Bush
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
 ; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
 ; NUMBER OF SEQUENCES: 7310
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
 ; STREET: 100 Beaver Street
 ; CITY: Waltham
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02354
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: CD-ROM ISO9660
 ; COMPUTER: PC
 ; OPERATING SYSTEM: <Unknown>
 ; SOFTWARE: ASCII
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/107,532A
 ; FILING DATE: 30-Jun-1998
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 60/085,598
 ; FILING DATE: 14 May 1998
 ; APPLICATION NUMBER: 60/051571
 ; FILING DATE: July 2, 1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Ariniello, Pamela Deneke
 ; REGISTRATION NUMBER: 40,489
 ; REFERENCE/DOCKET NUMBER: GTC-012
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (781)893-5007
 ; TELEFAX: (781)893-8277
 ; INFORMATION FOR SEQ ID NO: 6771:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 63 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; HYPOTHETICAL: YES
 ; ORIGINAL SOURCE:
 ; ORGANISM: Enterococcus faecium
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (B) LOCATION 1...63
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 6771:
 US-09-107-532A-6771

Query Match 63.5%; Score 33; DB 4; Length 63;
 Best Local Similarity 45.5%; Pred. No. 8.5;
 Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVVPXGXHS 11
 Db 5 EEMAFGDHYN 15

RESULT 7

US-08-961-083-182

; Sequence 182, Application US/08961083
 ; Patent No. 6159469

; GENERAL INFORMATION:
 ; APPLICANT: Choi et. al.
 ; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
 ; NUMBER OF SEQUENCES: 452
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850

; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33
 ; OPERATING SYSTEM: MSDOS version 6.2
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/961,083

; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:

; ATTORNEY/AGENT INFORMATION:
 ; NAME: Brookes, A. Anders
 ; REGISTRATION NUMBER: 36,373
 ; REFERENCE/DOCKET NUMBER: PB340P2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (301) 309-8504

; TELEFAX: (301) 309-8512
 ; INFORMATION FOR SEQ ID NO: 182:
 ; SEQUENCE CHARACTERISTICS:

; LENGTH: 447 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-961-083-182

Query Match 63.5%; Score 33; DB 3; Length 447;
 Best Local Similarity 62.5%; Pred. No. 74;

Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 3 VVPXGXHY 10

Db 178 IVPHGHHY 185

RESULT 8

US-09-536-784-182
 ; Sequence 182, Application US/09536784
 ; Patent No. 6573082

; GENERAL INFORMATION:

; APPLICANT: Choi et. al.
 ; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
 ; NUMBER OF SEQUENCES: 452
 ; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
 ; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2
 ; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/536,784
 ; FILING DATE: 30-Oct-1997

```
;
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 796 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-961-083-56

Query Match      63.5%; Score 33; DB 3; Length 796;
Best Local Similarity 62.5%; Pred. No. 1.4e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      3 VVPKXGXY 10
      :|||
Db      185 IVPKGDHY 192
```

```
RESULT 15
US-09-536-784-56
; Sequence 56, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/536,784
; FILING DATE: 30-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 796 amino acids
; TYPE: amino acid
```

```
;
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 56:
; US-09-536-784-56

Query Match      63.5%; Score 33; DB 4; Length 796;
Best Local Similarity 62.5%; Pred. No. 1.4e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      3 VVPKXGXY 10
      :|||
Db      185 IVPKGDHY 192

Search completed: December 22, 2003, 16:43:44
Job time : 10.1333 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:40:14 ; Search time 20.8667 Seconds
(without alignments)
98.451 Million cell updates/sec

Title: US-09-909-164-42

Perfect score: 52

Sequence: 1 BEVPGXGHYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 696363 seqs, 186758610 residues

Total number of hits satisfying chosen parameters: 696363

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:
1: /cgm2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgm2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgm2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgm2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgm2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgm2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgm2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgm2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgm2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgm2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgm2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgm2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgm2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgm2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgm2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgm2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgm2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgm2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------------|
| 1 | 34 | 65.4 | 381 | 12 | US-09-769-736-24 |
| 2 | 34 | 65.4 | 715 | 11 | US-09-252-088-16 |
| 3 | 34 | 65.4 | 715 | 12 | US-10-340-792-16 |
| 4 | 34 | 65.4 | 793 | 11 | US-09-252-088-15 |
| 5 | 34 | 65.4 | 793 | 12 | US-10-340-792-15 |
| 6 | 34 | 65.4 | 822 | 12 | US-09-769-736-18 |
| 7 | 33 | 63.5 | 447 | 9 | US-09-765-272-182 |
| 8 | 33 | 63.5 | 484 | 11 | US-09-769-787-38 |
| 9 | 33 | 63.5 | 484 | 12 | US-09-769-744A-24 |
| 10 | 33 | 63.5 | 763 | 9 | US-09-765-272-66 |
| 11 | 33 | 63.5 | 796 | 9 | US-09-765-272-56 |
| 12 | 33 | 63.5 | 826 | 11 | US-09-769-787-194 |
| 13 | 33 | 63.5 | 838 | 11 | US-09-884-465A-8 |
| 14 | 33 | 63.5 | 840 | 11 | US-09-884-465A-7 |
| 15 | 33 | 63.5 | 840 | 11 | US-09-884-465A-10 |

Sequence 6, Appli
Sequence 2486, A
Sequence 24, Appl
Sequence 24, Appl
Sequence 11332, A
Sequence 164, Appl
Sequence 36, Appl
Sequence 38, Appl
Sequence 22, Appl
Sequence 36, Appl
Sequence 343, App
Sequence 292, App
Sequence 13210, A
Sequence 104, App
Sequence 31, Appl
Sequence 8, Appl
Sequence 10416, A
Sequence 27, Appl
Sequence 10, Appl
Sequence 4919, Ap
Sequence 26, Appl
Sequence 2, Appl
Sequence 61, Appl
Sequence 2, Appl
Sequence 7521, Ap
Sequence 4, Appl
Sequence 32, Appl
Sequence 32, Appl
Sequence 67, Appl
Sequence 2, Appl

US-09-884-465A-6
US-10-029-386-12486
US-10-038-307-24
US-10-201-292-24
US-10-156-761-11332
US-09-934-455-164
US-10-302-267-36
US-10-171-404A-38
US-10-171-404A-36
US-10-176-847-22
US-10-205-823-343
US-09-881-752A-292
US-10-156-761-13210
US-10-279-029-104
US-10-219-810-31
US-10-177-293-8
US-09-815-242-10416
US-10-298-638-10
US-10-298-638-10
US-09-738-626-4919
US-09-820-843A-26
US-09-942-447-2
US-10-391-399-61
US-10-319-762-2
US-10-032-585-7521
US-10-122-067-4
US-10-262-794A-32
US-10-242-056-32
US-10-096-534-67
US-10-122-067-2

ALIGNMENTS

RESULT 1

US-09-769-736-24
; Sequence 24, Application US/09769736
; Publication No. US20030138775A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Le Page, Richard WF
; APPLICANT: Wells, Jeremy M
; APPLICANT: Hanniffy, Sean B
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21089wo
; CURRENT APPLICATION NUMBER: US/09/769,736
; CURRENT FILING DATE: 2003-02-14
; PRIOR APPLICATION NUMBER: GB 9816335.5
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: US 60/125163
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
; LENGTH: 381
; TYPE: PRT
; ORGANISM: Streptococcus agalactiae
US-09-769-736-24

Query Match 65.4%; Score 34; DB 12; Length 381;
Best Local Similarity 75.0%; Pred. No. 98;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 VVPXGXHY 10

Db 350 VVPHGDHY 357

RESULT 2

US-09-252-088-16
; Sequence 16, Application US/09252088
; Publication No. US20030031682A1

; GENERAL INFORMATION:
 ; APPLICANT: BRODEUR, Bernard R.
 ; APPLICANT: RIOUX, Clement
 ; APPLICANT: BOYER, Martine
 ; APPLICANT: CHARLEBOIS, Isabelle
 ; APPLICANT: HAMEL, Jose
 ; APPLICANT: MARTIN, Denis
 ; TITLE OF INVENTION: NOVEL GROUP B STREPTOCOCCUS ANTIGENS
 ; FILE REFERENCE: 8331-9002
 ; CURRENT APPLICATION NUMBER: US/09/252,088
 ; CURRENT FILING DATE: 1999-02-18
 ; EARLIER APPLICATION NUMBER: US/60/075,425
 ; EARLIER FILING DATE: 1998-02-20
 ; NUMBER OF SEQ ID NOS: 44
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 16
 ; LENGTH: 715
 ; TYPE: PRT
 ; ORGANISM: group B streptococcus
 US-09-252-088-16

Query Match 65.4%; Score 34; DB 11; Length 715;
 Best Local Similarity 75.0%; Pred. No. 1.9e+02;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 VVPXGXY 10
 |||||
 Db 243 VVPHGDHY 250

RESULT 3
 US-10-340-792-16
 ; Sequence 16, Application US/10340792
 ; Publication No. US20030228323A1
 ; GENERAL INFORMATION:
 ; APPLICANT: BRODEUR, Bernard R.
 ; APPLICANT: RIOUX, Clement
 ; APPLICANT: BOYER, Martine
 ; APPLICANT: CHARLEBOIS, Isabelle
 ; APPLICANT: HAMEL, Josee
 ; APPLICANT: MARTIN, Denis
 ; TITLE OF INVENTION: NOVEL GROUP B STREPTOCOCCUS ANTIGENS
 ; FILE REFERENCE: 8331-9002
 ; CURRENT APPLICATION NUMBER: US/10/340,792
 ; CURRENT FILING DATE: 2003-01-13
 ; PRIOR APPLICATION NUMBER: US/09/252,088B
 ; PRIOR FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US/60/075,425
 ; PRIOR FILING DATE: 1998-02-20
 ; NUMBER OF SEQ ID NOS: 45
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 16
 ; LENGTH: 715
 ; TYPE: PRT
 ; ORGANISM: group B streptococcus
 US-10-340-792-16

Query Match 65.4%; Score 34; DB 12; Length 715;
 Best Local Similarity 75.0%; Pred. No. 1.9e+02;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 VVPXGXY 10
 |||||
 Db 243 VVPHGDHY 250

RESULT 4
 US-09-252-088-15
 ; Sequence 15, Application US/09252088
 ; Publication No. US20030031682A1
 ; GENERAL INFORMATION:
 ; APPLICANT: BRODEUR, Bernard R.
 ; APPLICANT: RIOUX, Clement

; APPLICANT: BOYER, Martine
 ; APPLICANT: CHARLEBOIS, Isabelle
 ; APPLICANT: HAMEL, Jose
 ; APPLICANT: MARTIN, Denis
 ; TITLE OF INVENTION: NOVEL GROUP B STREPTOCOCCUS ANTIGENS
 ; FILE REFERENCE: 8331-9002
 ; CURRENT APPLICATION NUMBER: US/09/252,088
 ; CURRENT FILING DATE: 1999-02-18
 ; EARLIER APPLICATION NUMBER: US/60/075,425
 ; EARLIER FILING DATE: 1998-02-20
 ; NUMBER OF SEQ ID NOS: 44
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 15
 ; LENGTH: 793
 ; TYPE: PRT
 ; ORGANISM: group B streptococcus
 US-09-252-088-15

Query Match 65.4%; Score 34; DB 11; Length 793;
 Best Local Similarity 75.0%; Pred. No. 2.1e+02;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 VVPXGXY 10
 |||||
 Db 321 VVPHGDHY 328

RESULT 5
 US-10-340-792-15
 ; Sequence 15, Application US/10340792
 ; Publication No. US20030228323A1
 ; GENERAL INFORMATION:
 ; APPLICANT: BRODEUR, Bernard R.
 ; APPLICANT: RIOUX, Clement
 ; APPLICANT: BOYER, Martine
 ; APPLICANT: CHARLEBOIS, Isabelle
 ; APPLICANT: HAMEL, Josee
 ; APPLICANT: MARTIN, Denis
 ; TITLE OF INVENTION: NOVEL GROUP B STREPTOCOCCUS ANTIGENS
 ; FILE REFERENCE: 8331-9002
 ; CURRENT APPLICATION NUMBER: US/10/340,792
 ; CURRENT FILING DATE: 2003-01-13
 ; PRIOR APPLICATION NUMBER: US/09/252,088B
 ; PRIOR FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US/60/075,425
 ; PRIOR FILING DATE: 1998-02-20
 ; NUMBER OF SEQ ID NOS: 45
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 15
 ; LENGTH: 793
 ; TYPE: PRT
 ; ORGANISM: group B streptococcus
 US-10-340-792-15

Query Match 65.4%; Score 34; DB 12; Length 793;
 Best Local Similarity 75.0%; Pred. No. 2.1e+02;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 VVPXGXY 10
 |||||
 Db 321 VVPHGDHY 328

RESULT 6
 US-09-769-736-18
 ; Sequence 18, Application US/09769736
 ; Publication No. US20030138775A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Microbial Technics Limited
 ; APPLICANT: Le Page, Richard WF
 ; APPLICANT: Wells, Jeremy M
 ; APPLICANT: Hanniffy, Sean B
 ; TITLE OF INVENTION: Proteins

```
; FILE REFERENCE: PWC/P21089wo
; CURRENT APPLICATION NUMBER: US/09/769,736
; CURRENT FILING DATE: 2003-02-14
; PRIOR APPLICATION NUMBER: GB 9816335.5
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: US 60/125163
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 822
; TYPE: PRT
; ORGANISM: Streptococcus agalactiae
US-09-769-736-18

Query Match 65.4%; Score 34; DB 12; Length 822;
Best Local Similarity 75.0%; Pred. No. 2.2e+02; Indels 0; Gaps 0;
Matches 6; Conservative 0; Mismatches 2;

Qy 3 VVPXGXHY 10
Db 350 VVPHGDHY 357

RESULT 7
US-09-765-272-182
; Sequence 182, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340B2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 182:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 447 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 182:
US-09-765-272-182

Query Match 63.5%; Score 33; DB 9; Length 447;
Best Local Similarity 62.5%; Pred. No. 1.8e+02; Indels 0; Gaps 0;
Matches 5; Conservative 1; Mismatches 2;

Qy 3 VVPXGXHY 10
Db 350 VVPHGDHY 357

US-09-769-787-38
; Sequence 38, Application US/09769787
; Publication No. US20030091577A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21129WO
; CURRENT APPLICATION NUMBER: US/09/769,787
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 38
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-769-787-38

Query Match 63.5%; Score 33; DB 11; Length 484;
Best Local Similarity 62.5%; Pred. No. 2e+02; Indels 0; Gaps 0;
Matches 5; Conservative 1; Mismatches 2;

Qy 3 VVPXGXHY 10
Db 202 IVPHGGHY 209

RESULT 9
US-09-769-744A-24
; Sequence 24, Application US/09769744A
; Publication No. US20030134407A1
; GENERAL INFORMATION:
; APPLICANT: Le Page, Richard WF
; APPLICANT: Wells, Jeremy M
; APPLICANT: Hanniffy, Sean B
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21122WO
; CURRENT APPLICATION NUMBER: US/09/769,744A
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: PCT/GB99/02452
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: GB 9816336.3
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: US 60/125329
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-769-744A-24

Query Match 63.5%; Score 33; DB 12; Length 484;
Best Local Similarity 62.5%; Pred. No. 2e+02; Indels 0; Gaps 0;
Matches 5; Conservative 1; Mismatches 2;

Qy 3 VVPXGXHY 10
Db 202 IVPHGGHY 209
```

RESULT 10
US-09-765-272-66
; Sequence 66, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 763 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 66:
US-09-765-272-66
Query Match 63.5%; Score 33; DB 9; Length 763;
Best Local Similarity 62.5%; Pred. No. 3.2e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 3 VVPXGXHY 10
Db 184 IVPHGDHY 191
RESULT 11
US-09-765-272-56
; Sequence 56, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 796 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 56:
US-09-765-272-56
Query Match 63.5%; Score 33; DB 9; Length 796;
Best Local Similarity 62.5%; Pred. No. 3.4e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 3 VVPXGXHY 10
Db 185 IVPHGDHY 192
RESULT 12
US-09-769-787-194
; Sequence 194, Application US/09769787
; Publication No. US20030091577A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21129WO
; CURRENT APPLICATION NUMBER: US/09/769,787
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 194
; LENGTH: 826
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-769-787-194
Query Match 63.5%; Score 33; DB 11; Length 826;
Best Local Similarity 62.5%; Pred. No. 3.5e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 3 VVPXGXHY 10
Db 204 IVPHGDHY 211
RESULT 13
US-09-884-465A-8
; Sequence 8, Application US/09884465A
; Publication No. US2003007293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee

```

; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR FILING DATE: 60/212,683
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 838
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-884-465A-8

```

```

Query Match      63.5%; Score 33; DB 11; Length 838;
Best Local Similarity 62.5%; Pred. No. 3.6e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy      3 VVPXGXHY 10
       :|||
Db      203 IVPXGDHY 210

```

RESULT 14

```

US-09-884-465A-7
; Sequence 7, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR FILING DATE: 60/212,683
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 840
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-884-465A-7

```

```

Query Match      63.5%; Score 33; DB 11; Length 840;
Best Local Similarity 62.5%; Pred. No. 3.6e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy      3 VVPXGXHY 10
       :|||
Db      205 IVPXGDHY 212

```

RESULT 15

```

US-09-884-465A-10
; Sequence 10, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens

```

```

; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR FILING DATE: 60/212,683
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 840
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-884-465A-10

```

```

Query Match      63.5%; Score 33; DB 11; Length 840;
Best Local Similarity 62.5%; Pred. No. 3.6e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy      3 VVPXGXHY 10
       :|||
Db      182 IVPXGXHY 189

```

```

Search completed: December 22, 2003, 17:32:44
Job time : 21.9333 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 17:24:36 ; Search time 9.06667 Seconds
(without alignments)
116.675 Million cell updates/sec

Title: US-09-909-164-42
Perfect score: 52
Sequence: 1 ERVVPXGXHY 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----------|--------------------|
| 1 | 35 | 67.3 | 233 | 2 T02590 | DNA binding protei |
| 2 | 35 | 67.3 | 308 | 2 A72207 | ftsH proteinase ac |
| 3 | 35 | 67.3 | 360 | 2 E69086 | cell division prot |
| 4 | 35 | 67.3 | 1028 | 2 AF3286 | ATP-dependent DNA |
| 5 | 35 | 67.3 | 1396 | 2 S36851 | L-shaped tail fibe |
| 6 | 34 | 65.4 | 102 | 2 A42452 | Vi protein - tobac |
| 7 | 34 | 65.4 | 126 | 2 A25905 | tachykinin B precu |
| 8 | 34 | 65.4 | 279 | 2 C75538 | hypothetical prote |
| 9 | 34 | 65.4 | 495 | 2 T28717 | hypothetical prote |
| 10 | 34 | 65.4 | 534 | 2 A69284 | coenzyme F420-quin |
| 11 | 34 | 65.4 | 822 | 2 T46758 | hypothetical 92.4K |
| 12 | 34 | 65.4 | 1057 | 2 F89892 | carbamoyl-phosphat |
| 13 | 33 | 63.5 | 46 | 2 E97985 | hypothetical prote |
| 14 | 33 | 63.5 | 124 | 1 VKL051 | trans-regulatory s |
| 15 | 33 | 63.5 | 156 | 2 D82618 | conserved hypotet |
| 16 | 33 | 63.5 | 252 | 2 AE2001 | hypothetical prote |
| 17 | 33 | 63.5 | 412 | 2 A48702 | 2-methyl-branched- |
| 18 | 33 | 63.5 | 460 | 2 S69046 | hypothetical prote |
| 19 | 33 | 63.5 | 510 | 2 G86430 | T518.1 protein - A |
| 20 | 33 | 63.5 | 743 | 2 S38143 | hypothetical prote |
| 21 | 33 | 63.5 | 802 | 2 C95136 | conserved domain p |
| 22 | 33 | 63.5 | 819 | 2 B95136 | conserved domain p |
| 23 | 33 | 63.5 | 828 | 2 B98004 | hypothetical prote |
| 24 | 33 | 63.5 | 839 | 2 G95115 | conserved hypotet |
| 25 | 33 | 63.5 | 853 | 2 C97985 | hypothetical prote |
| 26 | 33 | 63.5 | 855 | 2 D98004 | histidine Motif-Co |
| 27 | 33 | 63.5 | 1039 | 2 H95115 | conserved hypotet |
| 28 | 33 | 63.5 | 1039 | 2 D97985 | hypothetical prote |
| 29 | 32 | 61.5 | 162 | 2 A70939 | probable monoxigen |

| | | | | | |
|----|----|------|-----|----------|---------------------|
| 30 | 32 | 61.5 | 225 | 2 S57810 | hypothetical prote |
| 31 | 32 | 61.5 | 286 | 2 C86169 | prohibitin 2 limpo |
| 32 | 32 | 61.5 | 311 | 2 H69194 | GMP synthetase, su |
| 33 | 32 | 61.5 | 322 | 2 AF3211 | hypothetical prote |
| 34 | 32 | 61.5 | 364 | 2 S47540 | fructose-bisphosph |
| 35 | 32 | 61.5 | 369 | 2 T17267 | hypothetical prote |
| 36 | 32 | 61.5 | 374 | 2 G69119 | probable aspartate |
| 37 | 32 | 61.5 | 425 | 2 T24111 | hypothetical prote |
| 38 | 32 | 61.5 | 454 | 2 G96568 | probable non-photo |
| 39 | 32 | 61.5 | 519 | 2 G84598 | probable bZIP tran |
| 40 | 32 | 61.5 | 536 | 2 AI3544 | aldenhyde dehydroge |
| 41 | 32 | 61.5 | 627 | 2 A69663 | DNA mismatch repai |
| 42 | 32 | 61.5 | 696 | 2 A91247 | phage transposase |
| 43 | 32 | 61.5 | 701 | 2 S61239 | hypothetical prote |
| 44 | 32 | 61.5 | 716 | 1 JC5061 | macrophage-stimula |
| 45 | 32 | 61.5 | 738 | 2 A48246 | ethylene-response |

ALIGNMENTS

RESULT 1

T02590

DNA binding protein EREBP-2 - common tobacco

C/Species: Nicotiana tabacum (common tobacco)

C/Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 21-Jul-2000

C/Accession: T02590

R/Ohme-Takagi, M.; Shinshi, H.

Plant Cell 7, 173-182, 1995

A/Title: Ethylene-inducible DNA binding proteins that interact with an ethylene respon

A/Reference number: Z14671; MUID:95276459; PMID:7756828

A/Accession: T02590

A/Status: preliminary; translated from GB/EMBL/DDBJ

A/Molecule type: mRNA

A/Residues: 1-233 <OHM>

A/Cross-references: EMBL:D38126; NID:g790362; PIDN:BAA07324.1; PID:gl208498

A/Experimental source: strain BY4; tissue-type leaf

Query Match 67.3%; Score 35; DB 2; Length 233;

Best Local Similarity 60.0%; Pred. No. 9.8;

Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 ERVVPXGXHY 10

Db 90 QAVVPKGRHY 99

RESULT 2

A72207

ftsH proteinase activity modulator HflK - Thermotoga maritima (strain MSB8)

C/Species: Thermotoga maritima

C/Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000

C/Accession: A72207

R/Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hick-
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D
C.M.

Nature 399, 323-329, 1999

A/Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome s

A/Reference number: A72200; MUID:99287316; PMID:10360571

A/Accession: A72207

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-308 <ARN>

A/Cross-references: GB:AE001819; GB:AE000512; NID:g4982396; PIDN:AAD36885.1; PID:g4982

A/Experimental source: strain MSB8

C/Genetics:

A/Gene: TM1822

C/Superfamily: erythrocyte band 7 integral membrane protein

Query Match 67.3%; Score 35; DB 2; Length 308;

Best Local Similarity 75.0%; Pred. No. 13;

Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHY 10
|||||
Db 41 VVPXGIRY 48

RESULT 3

E69086
cell division protein - Methanobacterium thermoautotrophicum (strain Delta H)
C;Species: Methanobacterium thermoautotrophicum
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 29-Sep-1999
C;Accession: E69086
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
Qiu, D.; Spadafora, R.; Vicalire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwan, N.;
Ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
A;Reference number: A69000; MUID:98037514; PMID:19371463
A;Accession: E69086
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-360 <MTH>
A;Cross-references: GB:AE000923; GB:AE000666; NID:92622766; PIDN:AA886115.1; PID:9262277
A;Experimental source: strain Delta H
C;Genetics:
A;Gene: MTH1642
C;Superfamily: cell division protein MJ0174

Query Match 67.3%; Score 35; DB 2; Length 360;
Best Local Similarity 45.5%; Pred. No. 16;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 BEVVPXGXHY 11
|:|:|:|:|:|:
Db 98 EDLVPMSHHT 108

RESULT 4

AF3286
ATP-dependent DNA helicase BMEI0275 [imported] - Brucella melitensis (strain 16M)
C;Species: Brucella melitensis
C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C;Accession: AF3286
R;DeVechio, V.G.; Kaputral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
M.; Mazur, M.; Goltzman, B.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A;Title: The genome sequence of the facultative intracellular pathogen Brucella melitend
A;Reference number: AD3252; PMID:1175668
A;Accession: AF3286
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1028 <KUR>
A;Cross-references: GB:AE008917; PIDN:AA151457.1; PID:gl7982167; GSPDB:GN00190
A;Experimental source: strain 16M
C;Genetics:
A;Gene: BMEI0275
A;Map position: I

Query Match 67.3%; Score 35; DB 2; Length 1028;
Best Local Similarity 54.5%; Pred. No. 49;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXHY 11
|:|:|:|:|:|:
Db 76 EKIVPPGARYS 86

RESULT 5

S36851
L-shaped tail fiber protein - phage T5
N;Alternate names: ltf protein
C;Species: phage T5
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 17-Nov-2000
C;Accession: S65934; S01984; S36851

R;Kallman, A.V.; Kulshin, V.E.; Shlyapnikov, M.G.; Keenzenko, V.N.; Kryukov, V.M.
PEBS Lett. 366, 46-48, 1995
A;Title: The nucleotide sequence of the bacteriophage T5 ltf gene.
A;Reference number: S65934; MUID:95309401; PMID:7789514
A;Accession: S65934
A;Molecule type: DNA
A;Residues: 1-1396 <KAL>
A;Cross-references: EMBL:X69460; NID:gl5415; PIDN:CAA49220.1; PID:gl5416
R;Kallman, A.V.; Kryukov, V.M.; Bayev, A.A.
Nucleic Acids Res. 16, 6230, 1988
A;Title: The nucleotide sequence of bacteriophage T5 DNA at the region between early a
A;Reference number: S01982; MUID:88289370; PMID:3267228
A;Accession: S01984
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 934-985, 'A', 987-1396 <KA2>
A;Cross-references: EMBL:X07559
C;Genetics:
A;Gene: ltf
C;Keywords: late protein; tail fiber

Query Match 67.3%; Score 35; DB 2; Length 1396;
Best Local Similarity 50.0%; Pred. No. 68;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 EVVPXGXHY 11
|:|:|:|:|:|:
Db 1360 KIIPAGDHYS 1369

RESULT 6

A42452
V1 protein - tobacco yellow dwarf virus (strain Australia)
C;Species: tobacco yellow dwarf virus
C;Date: 15-Jan-1993 #sequence_revision 15-Jan-1993 #text_change 08-Oct-1999
C;Accession: A42452
R;Morris, B.A.M.; Richardson, K.A.; Halsey, A.; Zhan, X.; Thomas, J.E.
Virology 187, 633-642, 1992
A;Title: The nucleotide sequence of the infectious cloned DNA component of tobacco yel
A;Reference number: A42452; MUID:92188538; PMID:1546458
A;Accession: A42452
A;Molecule type: DNA
A;Residues: 1-102 <MOR>
A;Cross-references: GB:M81103; NID:g335283; PIDN:AAA47947.1; PID:g335284

Query Match 65.4%; Score 34; DB 2; Length 102;
Best Local Similarity 60.0%; Pred. No. 6.5;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPXGXHY 11
|:|:|:|:|:|:
Db 7 QVVPXGINS 16

RESULT 7

A25905
tachykinin B precursor - bovine
N;Alternate names: neuromedin K
C;Species: Bos primigenius taurus (cattle)
C;Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 16-Jul-1999
C;Accession: A25905
R;Kotani, H.; Hoshimaru, M.; Nawa, H.; Nakanishi, S.
Proc. Natl. Acad. Sci. U.S.A. 83, 7074-7078, 1986
A;Title: Structure and gene organization of bovine neuromedin K precursor.
A;Reference number: A25905; MUID:86313713; PMID:3462746
A;Accession: A25905
A;Molecule type: DNA
A;Residues: 1-126 <KOT>
A;Cross-references: GB:M14351; NID:gl63587; PIDN:AAA30723.1; PID:gl63590
C;Superfamily: neurokinin B precursor
F;1-26/Domain: signal sequence #status predicted <SIG>
F;27-126/Product: protachykinin B #status predicted <MAT>

Query Match 65.4%; Score 34; DB 2; Length 126;
 Best Local Similarity 66.7%; Pred. No. 8.2;
 Matches 6; Conservative 1; Mismatches 0; Gaps 0;

Qy 1 BEVVPXGXH 9
 Db 28 EQVPGGHH 36
 ||| |||
 ||| |||

RESULT 8
 C75538
 hypothetical protein - Deinococcus radiodurans (strain R1)
 C:Species: Deinococcus radiodurans
 C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 28-Jul-2000
 C:Accession: C75538
 R:White, O.; Eisen, J.A.; Heideberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
 S.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
 Science 286, 1571-1577, 1999
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A:Reference number: A75250; MUID:20036896; PMID:10567266
 A:Accession: C75538
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-279 <WHI>
 A:Cross-references: GB:AE001889; GB:AE000513; NID:G6457944; PIDN:AAF09867.1; PID:G645795
 A:Experimental source: strain R1
 C:Genetics:
 A:Gene: DR0271
 A:Map position: 1
 C:Superfamily: Deinococcus radiodurans hypothetical protein DR0271

Query Match 65.4%; Score 34; DB 2; Length 279;
 Best Local Similarity 75.0%; Pred. No. 19;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 4 VPXGXHYS 11
 Db 100 VPLGRHYS 107
 ||| ||| |||
 ||| ||| |||

RESULT 9
 T28717
 hypothetical protein F10D2.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T28717
 R:Graves, T.; Wohldmann, P.; Gillam, B.
 submitted to the EMBL Data Library, September 1997
 A:Description: The sequence of C. elegans cosmid F10D2.
 A:Reference number: Z20515
 A:Accession: T28717
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-495 <GRA>
 A:Cross-references: EMBL:AF022972; PIDN:AAC48234.1; GSPDB:GN00023; CESP:F10D2.3
 A:Experimental source: strain Bristol N2; clone F10D2
 C:Genetics:
 A:Gene: CESP:F10D2.3
 A:Map position: 5
 A:Introns: 37/2; 90/1; 113/3; 356/3; 380/3; 428/2

Query Match 65.4%; Score 34; DB 2; Length 495;
 Best Local Similarity 50.0%; Pred. No. 36;
 Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVVPXGXHY 10
 Db 218 ENIVPTGRKH 227
 ||| ||| |||
 ||| ||| |||

RESULT 10
 A69284

coenzyme F420-quinone oxidoreductase (EC 1.6.5.-) 56K chain - Archaeoglobus fulgidus
 N:Alternate names: sarcosine oxidase alpha chain soxA homolog
 C:Species: Archaeoglobus fulgidus
 C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 19-May-2000
 C:Accession: A69284; S45665
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dode
 ; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.P.; McDonald, L.
 Nature 390, 364-370, 1997
 A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes,
 Smith, H.O.; Woese, C.R.; Venter, J.C.
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archa
 A:Reference number: A69250; MUID:98049343; PMID:9389475
 A:Accession: A69284
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-534 <KLE>
 A:Cross-references: GB:AE001086; GB:AE000782; NID:G2689409; PIDN:AAB90959.1; PID:G2650
 R:Kunov, J.; Linder, D.; Stetter, K.O.; Thauer, R.K.
 Eur. J. Biochem. 223, 503-511, 1994
 A:Title: F(420)H(2): quinone oxidoreductase from Archaeoglobus fulgidus. Characterizat
 A:Reference number: S45665; MUID:94333340; PMID:8055920
 A:Accession: S45665
 A:Molecule type: protein
 A:Residues: 2 'X', 4 'X', 6-7 'X', 9-10 'XX', 13-14 <KUN>
 A:Note: the authors could not distinguish between glutamate and cysteine for residues
 C:Superfamily: glutamate synthase small chain
 C:Keywords: oxidoreductase

Query Match 65.4%; Score 34; DB 2; Length 534;
 Best Local Similarity 45.5%; Pred. No. 39;
 Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 BEVVPXGXHYS 11
 Db 119 DKVFPFGSHVT 129
 ||| ||| |||
 ||| ||| |||

RESULT 11
 T46758
 hypothetical 92.4K protein - Streptococcus agalactiae
 C:Species: Streptococcus agalactiae
 C>Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 21-Jul-2000
 C:Accession: T46758
 R:Spellerberg, B.; Rozdinski, E.; Martin, S.; Weber-Heynemann, J.; Schnitzler, N.; Lu
 Infect. Immun. 67, 871-878, 1999
 A:Title: Lmb, a protein with similarities to the Iral adhesin family, mediates attachm
 A:Reference number: Z24091; MUID:99115568; PMID:9916102
 A:Accession: T46758
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-822 <SPE>
 A:Cross-references: EMBL:AF062533; NID:G4249622; PIDN:AAD13797.1; PID:G4249624
 A:Experimental source: strain R268
 C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 65.4%; Score 34; DB 2; Length 822;
 Best Local Similarity 75.0%; Pred. No. 62;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 VVPXGXHY 10
 Db 350 VVPFGDHY 357
 ||| ||| |||
 ||| ||| |||

RESULT 12
 F89892
 carbamoyl-phosphate synthase large chain [imported] - Staphylococcus aureus (strain N3
 C:Species: Staphylococcus aureus
 C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
 C:Accession: F89892
 R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Og
 ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, R.; Kaito, C.; Sekimizu, K.

C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
 Lancet 357, 1225-1240, 2001
 A;Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.
 A;Reference number: A89758; MUID:21311952; PMID:11418146
 A;Accession: F89992
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-1057 <KUR>
 A;Cross-references: GB:BA000018; PID:g13701002; PIDN:BA842298.1; GSPDB:GN00149
 A;Experimental source: strain N315
 C;Genetics:
 A;Gene: pVRAB
 C;Superfamily: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain; biotin

Query Match 65.4%; Score 34; DB 2; Length 1057;
 Best Local Similarity 60.0%; Pred. No. 82;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EVVPXGXHY 11
 ||| |||
 Db 190 EIVSNGLHY 199

RESULT 13

E97985
 hypothetical protein spr0909 [imported] - Streptococcus pneumoniae (strain R6)
 C;Species: Streptococcus pneumoniae
 C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
 C;Accession: E97985
 R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E
 e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
 y, P.; Sun, P.M.; Winkler, M.E.
 J. Bacteriol. 183, 5709-5717, 2001
 A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.
 A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
 A;Reference number: A97872; MUID:21429245; PMID:11544234
 A;Accession: E97985
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-46 <KUR>
 A;Cross-references: GB:AE007317; PIDN:AAK99713.1; PID:g15458516; GSPDB:GN00174
 C;Genetics:
 A;Gene: spr0909

Query Match 63.5%; Score 33; DB 2; Length 46;
 Best Local Similarity 62.5%; Pred. No. 4.5;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXHY 10
 ||| |||
 Db 21 IVPHGHHY 28

RESULT 14

VKLJSI
 trans-regulatory splicing protein - simian immunodeficiency virus SIVcpz
 N;Alternate names: anti-repression trans-activator; art protein; rev protein; trs protei
 C;Species: simian immunodeficiency virus SIVcpz
 A;Note: host Pan troglodytes (chimpanzee)
 C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 16-Jul-1999
 C;Accession: S09988
 R;Huet, T.; Cheynier, R.; Meyerhans, A.; Roelants, G.; Wain-Hobson, S.
 Nature 345, 356-359, 1990
 A;Title: Genetic organization of a chimpanzee lentivirus related to HIV-1.
 A;Reference number: S09983; MUID:90259077; PMID:2188136
 A;Accession: S09988
 A;Status: nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-124 <HUR>
 A;Cross-references: EMBL:X52154; NID:G58866; PIDN:CAA36405.1; PID:g763085
 C;Genetics:
 A;Gene: rev; trs; art
 A;Introns: 27/1

C;Superfamily: AIDS trans-regulatory splicing protein
 C;Keywords: AIDS; immunodeficiency; splicing protein; transcription regulation

Query Match 63.5%; Score 33; DB 1; Length 124;
 Best Local Similarity 60.0%; Pred. No. 13;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 EVVPXGXHY 11
 ||| |||
 Db 107 ETVPAGGNY 116

RESULT 15

D82618
 conserved hypothetical protein XF1950 [imported] - Xylella fastidiosa (strain 9a5c)
 C;Species: Xylella fastidiosa
 C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
 C;Accession: D82618
 R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequ
 Nature 406, 151-157, 2000
 A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A;Reference number: A82515; MUID:20365717; PMID:10910347
 A;Note: for a complete list of authors see reference number A59328 below
 A;Accession: D82618
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-156 <SIM>
 A;Cross-references: GB:AE004014; GB:AE003849; NID:g9107044; PIDN:AAF84752.1; GSPDB:GN0
 A;Experimental source: strain 9a5c
 R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.;
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carre
 as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
 submitted to GenBank, June 2000
 A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fro
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuranae, E.B.; Lai
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins,
 A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, B.C.; Miyaki, C.Y
 P.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.B.; de Sa, R.G.; Santelli, R.V.; Sawas
 A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silve
 M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.;
 A;References: number: A59328
 A;Contents: annotation
 C;Genetics:
 A;Gene: XF1950

Query Match 63.5%; Score 33; DB 2; Length 156;
 Best Local Similarity 55.6%; Pred. No. 17;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EVVPXGXH 9
 ||| |||
 Db 119 BEILPQGVH 127

Search completed: December 22, 2003, 17:44:59

Job time : 10.0667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:43:51 ; Search time 4.6 Seconds
(without alignments)
112.455 Million cell updates/sec

Title: US-09-909-164-42

Perfect score: 52

Sequence: 1 EEVVPXGXHS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|--------------|--------------------|
| 1 | 36 | 69.2 | 426 | 1 AROA_VIBPA | Q87qx9 vibrio para |
| 2 | 35 | 67.3 | 1058 | 1 CARB_FUSNN | Q8986 fusobacteri |
| 3 | 35 | 67.3 | 1396 | 1 VITF_BPT5 | P13390 bacterioph |
| 4 | 34 | 65.4 | 102 | 1 Y11K_TYDVA | P31619 tobacco yel |
| 5 | 34 | 65.4 | 126 | 1 TKNK_BOVIN | P08858 bos taurus |
| 6 | 34 | 65.4 | 1057 | 1 CARB_STAAM | Q99ur5 staphylococ |
| 7 | 34 | 65.4 | 1057 | 1 CARB_STAAM | P58940 staphylococ |
| 8 | 33 | 63.5 | 124 | 1 REV_SIVCZ | P17280 chimpanzee |
| 9 | 33 | 63.5 | 460 | 1 UMEI_YEAST | Q03010 saccharomyc |
| 10 | 33 | 63.5 | 578 | 1 MDLB_BUCBP | Q89a96 buchnera ap |
| 11 | 33 | 63.5 | 743 | 1 YK47_YEAST | P36148 saccharomyc |
| 12 | 32 | 61.5 | 308 | 1 GAAB_METH | O26806 methanobact |
| 13 | 32 | 61.5 | 363 | 1 ALFB_SHEEP | P52210 ovine arter |
| 14 | 32 | 61.5 | 573 | 1 SUOX_DROME | Q9vwp4 drosophila |
| 15 | 32 | 61.5 | 627 | 1 MUTL_BACSU | P49850 bacillus su |
| 16 | 32 | 61.5 | 735 | 1 ETR1_BRAOL | O49230 brassica ol |
| 17 | 32 | 61.5 | 738 | 1 ETR1_ARATH | P49333 arabidopsis |
| 18 | 32 | 61.5 | 906 | 1 CRNC_MOUSE | P49452 mus musculu |
| 19 | 32 | 61.5 | 1374 | 1 YQ56_CAEEL | Q09541 caenorhabdi |
| 20 | 32 | 61.5 | 1378 | 1 RON_MOUSE | Q62190 mus musculu |
| 21 | 31 | 59.6 | 130 | 1 S205_RAT | P97885 rattus norv |
| 22 | 31 | 59.6 | 212 | 1 MSRA_VIBCH | O9kp30 vibrio chol |
| 23 | 31 | 59.6 | 319 | 1 YHAI_CRYPA | P10941 cryptoneutr |
| 24 | 31 | 59.6 | 363 | 1 ALFB_HUMAN | P05062 homo sapien |
| 25 | 31 | 59.6 | 363 | 1 ALFB_RABIT | P79226 oryctolagus |
| 26 | 31 | 59.6 | 387 | 1 THIK_ECOLI | Q8x8j4 escherichia |
| 27 | 31 | 59.6 | 387 | 1 THIK_ECOLI | P21151 escherichia |
| 28 | 31 | 59.6 | 387 | 1 THIK_SALTY | Q91616 salmonella |
| 29 | 31 | 59.6 | 426 | 1 SL51_YARLI | Q99158 yarrowia li |
| 30 | 31 | 59.6 | 427 | 1 TOLB_HAEIN | P44677 haemophilus |
| 31 | 31 | 59.6 | 441 | 1 YL15_VIBPA | P46231 vibrio para |
| 32 | 31 | 59.6 | 991 | 1 DHP1_SCHPO | P40848 schizosacch |
| 33 | 31 | 59.6 | 1188 | 1 KPBA_CAEEL | P34335 caenorhabdi |

| | | | | | |
|----|------|------|------|--------------|--------------------|
| 34 | 31 | 59.6 | 1410 | 1 PDRB_YEAST | P40550 saccharomyc |
| 35 | 31 | 59.6 | 1499 | 1 A10C_HUMAN | O60312 homo sapien |
| 36 | 31 | 59.6 | 3099 | 1 POLG_PEMVM | O56075 p genome po |
| 37 | 30.5 | 58.7 | 268 | 1 IL1A_BOVIN | P08831 bos taurus |
| 38 | 30.5 | 58.7 | 268 | 1 IL1A_CAPIH | P79161 capra hircu |
| 39 | 30.5 | 58.7 | 268 | 1 IL1A_SHEEP | Q29579 ovine aries |
| 40 | 30 | 57.7 | 121 | 1 TKNK_HUMAN | Q9uhf0 homo sapien |
| 41 | 30 | 57.7 | 152 | 1 Y16D_BPT4 | P22917 bacterioph |
| 42 | 30 | 57.7 | 160 | 1 Y201_METJA | O27654 methanococc |
| 43 | 30 | 57.7 | 223 | 1 COAT_CTV36 | Q00686 citrus tris |
| 44 | 30 | 57.7 | 232 | 1 SCOA_HELPJ | Q92163 helicobacte |
| 45 | 30 | 57.7 | 232 | 1 SCOA_HELPJ | P56006 helicobacte |

ALIGNMENTS

RESULT 1
AROA_VIBPA
ID AROA_VIBPA STANDARD; PRT; 426 AA.
AC Q87OX9;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE 3-phosphohikimate 1-carboxyvinyltransferase (EC 2.5.1.19) (5-
enolpyruvylhikimate-3-phosphate synthase) (EPSP synthase) (EPSPS).
GN AROA OR VP1020.
OS Vibrio parahaemolyticus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=670;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RIMD 2210633 / Serotype O3:K6;
RX MEDLINE=22508454; PubMed=12620739;
RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
Iijima Y., Majima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
"Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
distinct from that of V. cholerae".
RT Lancet 361:743-749(2003).
RL
CC -1- CATALYTIC ACTIVITY: Phosphoenolpyruvate + 3-phosphohikimate =
phosphate + 5-O-(1-carboxyvinyl)-3-phosphohikimate.
CC -1- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
sixth step.
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: Belongs to the EPSP synthase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; AP005076; BAC59283.1; -
DR HAMAP; MF 00210; -; 1.
DR PROSITE; PS00104; EPSP SYNTHASE 1; 1.
DR PROSITE; PS00885; EPSP SYNTHASE 2; 1.
KW Aromatic amino acid biosynthesis; transferase; Complete proteome.
SQ SEQUENCE 426 AA; 46094 MW; 373939CC5BA1F70F CRC64;

Query Match 69.2%; Score 36; DB 1; Length 426;
Best Local Similarity 60.0%; Pred. No. 5.3;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 EEVVPXGXHY 10

Db 223 EFVIPAGQHY 232

YLIK TYDVA
ID YLIK TYDVA STANDARD; PRT; 102 AA.
AC P31619;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-OCT-1993 (Rel. 27, Last annotation update)
DE Hypothetical 11.2 kDa protein.
GN VI.
OS Tobacco yellow dwarf virus (strain Australia) (TYDV).
OS Viruses; ssDNA viruses; Geminiviridae; Mastrevirus.
OX NCBI_TaxID=31599;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92188538; PubMed=1546458;
RT Morris B.A.M., Richardson K.A., Haley A., Zhan X., Thomas J.E.;
RT "The nucleotide sequence of the infectious cloned DNA component of
RT tobacco yellow dwarf virus reveals features of geminiviruses
RT infecting monocotyledonous plants.";
RL Virology 187:633-642(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M81103; AAA47947.1; -
DR PIR; A42452; A42452.
DR InterPro; IPR002621; Gemin1_mov.
DR Pfam; PF01708; Gemin1_mov; 1.
DR Hypothetical protein.
KW Hypothetical protein.
SQ SEQUENCE 102 AA; 11178 MW; A40ECF1E0AF55B67 CRC64;
Query Match 65.4%; Score 34; DB 1; Length 102;
Best Local Similarity 60.0%; Pred. No. 3.2;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 EVVPXGXHYS 11
Db 7 QVPSGNGYS 16
RESULT 5
TKNK_BOVIN
ID TKNK_BOVIN STANDARD; PRT; 126 AA.
AC P08858;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurokinin B precursor (NKB) (Neuromedin K).
GN TAC3 OR NKNB.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86313713; PubMed=3462746;
RT Kotani H., Hoshimaru M., Nawa H., Nakanishi S.;
RT "Structure and gene organization of bovine neuromedin K precursor.";
RT Proc. Natl. Acad. Sci. U.S.A. 83:7074-7078(1986).
CC -1- FUNCTION: TACHYKININS ARE ACTIVE PEPTIDES WHICH EXCITE NEURONS,
CC EVOKE BEHAVIORAL RESPONSES, ARE POTENT VASODILATORS AND
CC SECRETAGOGUES, AND CONTRACT (DIRECTLY OR INDIRECTLY) MANY SMOOTH
CC MUSCLES.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE TACHYKININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M14351; AAA30723.1; -
DR EMBL; M14347; AAA30723.1; JOINED.
DR EMBL; M14348; AAA30723.1; JOINED.
DR EMBL; M14349; AAA30723.1; JOINED.
DR EMBL; M14350; AAA30723.1; JOINED.
DR PIR; A25905; A25905.
DR InterPro; IPR003635; Neurokinin.
DR InterPro; IPR002040; Tachykinin.
DR Pfam; PF03823; Neurokinin B; 1.
DR ProDom; PD020370; Neurokinin; 1.
DR PROSITE; PS00267; TACHYKININ; 1.
KW Tachykinin; Neuropeptide; Cleavage on pair of basic residues;
KW Amidation; Signal.
FT SIGNAL 1 20 POTENTIAL.
FT PROPEP 21 83
FT PEPTIDE 86 95 NEUROKININ B.
FT PROPEP 99 126
FT MOD_RES 95 95
SQ SEQUENCE 126 AA; 13871 MW; 446BF433498EC059 CRC64;
Query Match 65.4%; Score 34; DB 1; Length 126;
Best Local Similarity 66.7%; Pred. No. 4;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ERYVPXGXH 9
Db 28 EQVPGGGH 36
RESULT 6
CARB_STRAAM
ID CARB_STRAAM STANDARD; PRT; 1057 AA.
AC Q99UR5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
DE phosphate synthetase ammonia chain).
GN CARB OR PYRAB OR SAV1203 OR SA1046.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OS Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158878; 158879;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Mu50 / ATCC 700699, and N315;
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iino T., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus".
RL Lancet 357:1225-1240(2001).
CC -1- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
CC phosphate + L-glutamate + carbamoyl phosphate.
CC -1- COPACITOR: Binds 3 manganese ions per subunit (By similarity).
CC -1- PATHWAY: Arginine biosynthesis.
CC -1- SUBUNIT: Pyrimidine biosynthesis; first step.
CC -1- PROMOTES: Composed of two chains; the small (or glutamine) chain
CC promotes the hydrolysis of glutamine to ammonia, which is used by
CC the large (or ammonia) chain to synthesize carbamoyl phosphate (By
CC similarity).
CC -1- SIMILARITY: BELONGS TO THE CARB FAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AP003361; BAB57365.1; --
 EMBL; AP003132; BAB42298.1; --
 PIR; F89892; F89892.
 HSSP; P00968; 1CS0.
 HAMAP; MF 01210; --; 1.
 InterPro; IPR006275; CarA_L_glu.
 InterPro; IPR005483; CPhase_L.
 InterPro; IPR005479; CPhase_L_D2.
 InterPro; IPR005480; CPhase_L_D3.
 InterPro; IPR005481; CPhase_L_N.
 InterPro; IPR004362; MGS like.
 Pfam; PF00289; CPhase_L_Chain; 2.
 Pfam; PF02786; CPhase_L_D2; 2.
 Pfam; PF02787; CPhase_L_D3; 1.
 Pfam; PF02142; MGS; 1.
 PRINTS; PR00098; CPSASE.
 TIGRFAMS; TIGR01369; CPSaseII_lrg; 1.
 PROSITE; PS00866; CPSASE_1; 2.
 PROSITE; PS00867; CPSASE_2; 2.
 Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
 ATP-binding; Manganese; Complete proteome.
 FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
 FT DOMAIN 402 546 OLIGOMERIZATION DOMAIN.
 FT DOMAIN 547 929 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
 FT DOMAIN 930 1057 ALLOSTERIC DOMAIN.
 REPEAT 1 546
 REPEAT 547 1057
 NP_BIND 153 210 ATP (POTENTIAL).
 NP_BIND 302 352 ATP (POTENTIAL).
 METAL 284 284 MANGANESE 1 (BY SIMILARITY).
 METAL 298 298 MANGANESE 1 AND 2 (BY SIMILARITY).
 METAL 300 300 MANGANESE 2 (BY SIMILARITY).
 METAL 820 820 MANGANESE 3 (BY SIMILARITY).
 METAL 832 832 MANGANESE 3 (BY SIMILARITY).
 SQ SEQUENCE 1057 AA; 117171 MW; E3E179EF0591F0F8 CRC64;

Query Match 65.4%; Score 34; DB 1; Length 1057;
 Best Local Similarity 60.0%; Pred. No. 36;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 2 EVVPXGXHY 11
 Db 190 EIVSNGLHY 199

RESULT 7
 CARB STAAW STANDARD; PRT; 1057 AA.
 AC PS8940;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-phosphate synthetase ammonia chain).
 GN CARB OR PYRAB OR MW1086.
 OS Staphylococcus aureus (strain MW2).
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
 OX NCBI_TaxID=196620;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22040717; PubMed=12044378;
 RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A., Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L., Yamamoto K., Hiramatsu K.;

RT *Genome and virulence determinants of high virulence community-acquired MRSA.";
 RL Lancet 359:1819-1827(2002).
 CC -1- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP + phosphate + L-glutamate + carbamoyl phosphate.
 CC -1- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
 CC -1- PATHWAY: Arginine biosynthesis.
 CC -1- SUBUNIT: Pyrimidine biosynthesis; first step.
 CC promotes the hydrolysis of glutamine to ammonia, which is used by the large (or ammonia) chain to synthesize carbamoyl phosphate (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE CARB FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AP004825; BAB94951.1; --
 HAMAP; MF 01210; --; 1.
 InterPro; IPR006275; CarA_L_glu.
 InterPro; IPR005483; CPhase_L.
 InterPro; IPR005479; CPhase_L_D2.
 InterPro; IPR005480; CPhase_L_D3.
 InterPro; IPR005481; CPhase_L_N.
 InterPro; IPR004362; MGS like.
 Pfam; PF00289; CPhase_L_Chain; 2.
 Pfam; PF02786; CPhase_L_D2; 2.
 Pfam; PF02787; CPhase_L_D3; 1.
 Pfam; PF02142; MGS; 1.
 PRINTS; PR00098; CPSASE.
 TIGRFAMS; TIGR01369; CPSaseII_lrg; 1.
 PROSITE; PS00866; CPSASE_1; 2.
 PROSITE; PS00867; CPSASE_2; 2.
 Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
 ATP-binding; Manganese; Complete proteome.
 FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
 FT DOMAIN 402 546 OLIGOMERIZATION DOMAIN.
 FT DOMAIN 547 929 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
 FT DOMAIN 930 1057 ALLOSTERIC DOMAIN.
 REPEAT 1 546
 REPEAT 547 1057
 NP_BIND 153 210 ATP (POTENTIAL).
 NP_BIND 302 352 ATP (POTENTIAL).
 METAL 284 284 MANGANESE 1 (BY SIMILARITY).
 METAL 298 298 MANGANESE 1 AND 2 (BY SIMILARITY).
 METAL 300 300 MANGANESE 2 (BY SIMILARITY).
 METAL 820 820 MANGANESE 3 (BY SIMILARITY).
 METAL 832 832 MANGANESE 3 (BY SIMILARITY).
 SQ SEQUENCE 1057 AA; 117185 MW; D8E3B09F9BC6F152 CRC64;

Query Match 65.4%; Score 34; DB 1; Length 1057;
 Best Local Similarity 60.0%; Pred. No. 36;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 2 EVVPXGXHY 11
 Db 190 EIVSNGLHY 199

RESULT 8
 REV_SIVCZ
 ID REV_SIVCZ STANDARD; PRT; 124 AA.
 AC P17280;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE REV protein (Anti-repression transactivator protein) (ART/TRS).
 GN REV.

OS Chimpanzee immunodeficiency virus (SIV(cpz)) (CIV).
 OC Viruses; Retroviridae; Retroviridae; Lentivirus.
 OX NCBI_TaxID=11723;
 RN [1]
 RP SEQUENCE FROM N.A.
 EX MEDLINE=90259077; PubMed=2188136;
 RA Huet T., Cheynier R., Meyers A., Roelants G., Main-Hobson S.;
 RT "Genetic organization of a chimpanzee lentivirus related to HIV-1."
 RL Nature 345:356-359 (1990).
 CC -!- FUNCTION: REV APPEARS TO ACT POST-TRANSCRIPTIONALLY TO RELIEVE
 CC -!- NEGATIVE REPRESSION OF GAG AND ENV PRODUCTION.
 CC -!- SUBCELLULAR LOCATION: Nuclear; accumulates in the nucleoli.
 CC -!- PTM: PHOSPHORYLATION WHOSE STATE OF PHOSPHORYLATION IS MEDIATED
 CC BY A SPECIFIC SERINE KINASE ACTIVITY PRESENT IN THE NUCLEUS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X52154; CAJ36405.1; -;
 DR PIR; S09988; VKLJSI.
 DR HIV; X52154; REVSCPZ.
 DR InterPro; IPR000625; REV_protein.
 DR Pfam; PF00424; REV; 1.
 KW Transcription regulation; AIDS; Phosphorylation; Nuclear protein.
 SQ SEQUENCE 124 AA; 13701 MW; F5877D1BDP65A7B2 CRC64;

 Query Match 63.5%; Score 33; DB 1; Length 124;
 Best Local Similarity 60.0%; Pred. No. 6.5;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 OY 2 EVVPGXHY 11
 DB |||||
 107 ETVPGAGNYS 116

 RESULT 9
 ID UMEI YEAST STANDARD; PRT; 460 AA.
 AC Q03010; P87330;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Meiosis negative regulator UME1.
 GN UME1 OR WTM3 OR YPL139C OR LPI7C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A364A;
 RA Mallory M.J., Strich R.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c / AB972;
 RX MEDLINE=97313271; PubMed=9169875;
 RA Busey H., Storms R.K., Ahmed A., Albertmann K., Allen E., Ansoorge W.,
 RA Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,
 RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
 RA Chung E., Church C.M., Coster P., Davis K., Davis R.W.,
 RA Dietrich F.S., Delius H., DiPaolo T., Dubois E., Duesterhoeft A.,
 RA Duncan M., Floeth M., Portin N., Friesen J.D., Fritz C., Goffeau A.,
 RA Hall J., Hebling U., Heumann K., Hilbert H., Hallier L.,
 RA Hunnicke-Smith S., Hyman R., Johnston M., Kaiman S., Klein K.,
 RA Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,
 RA Marathe R., Messenguy F., Mews H.-W., Mirtipati S., Mostl D.,
 RA Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,

Petel F.X., Pohl T.M., Purnelle D., Schafer M., Schafke M.,
 RA Scherens B., Schramm S., Schroeder M., Sdicu A.M., Tettelin H.,
 RA Urrestarazu L.A., Ushinsky S., Vierendeels P., Viessers S., Voss H.,
 RA Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,
 RA Zhong W.M., Zollner A., Vo D.H., Hani J.;
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI."
 RL Nature 387:103-105 (1997).
 CC -!- FUNCTION: TRANSCRIPTIONAL MODULATOR WITH ROLES IN MEIOTIC
 CC REGULATION AND SILENCING. NEGATIVE REGULATOR OF MEIOSIS.
 CC -!- SIMILARITY: Contains 4 WD repeats.
 CC -!- SIMILARITY: STRONG, TO YEAST WTM1 AND WTM2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U10280; AAB40937.1; -;
 DR EMBL; U43703; AAB68221.1; -;
 DR PIR; S69046; S69046.
 DR TRANSFAC; T04309; -;
 DR SGD; S0006060; UME1.
 DR GO; GO:0005634; C:nucleus; IDA.
 DR GO; GO:0003714; P:transcription co-repressor activity; IDA.
 DR GO; GO:0040020; P:regulation of meiosis; IGI.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF00400; WD40; 3.
 DR SMART; SM00320; WD40; 4.
 DR PROSITE; PS00678; WD_REPEATS_1; FALSE NEG.
 DR PROSITE; PS50082; WD_REPEATS_2; FALSE NEG.
 DR PROSITE; PS50294; WD_REPEATS_REGION; FALSE NEG.
 KW Transcription regulation; Meiosis; Repeat; WD repeat.
 FT REPEAT 233 271 WD 1.
 FT REPEAT 276 316 WD 2.
 FT REPEAT 339 379 WD 3.
 FT REPEAT 411 451 WD 4.
 SQ SEQUENCE 460 AA; 51022 MW; AA6F60448B7BCBA9 CRC64;

 Query Match 63.5%; Score 33; DB 1; Length 460;
 Best Local Similarity 62.5%; Pred. No. 25;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 OY 3 VVPXGXHY 10
 DB :|||
 85 IVPLGLHY 92

 RESULT 10
 ID MDLB_BUCBP STANDARD; PRT; 578 AA.
 AC Q89A96;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Multidrug resistance-like ATP-binding protein mdlb.
 GN MDLB OR BHP424.
 OS Buchnera aphidicola (subsp. Baizongia pistaciae).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Buchnera.
 OX NCBI_TaxID=135842;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=2426901; PubMed=12522265;
 RA Van Ham R.C.H.J., Kamerbeek J., Palacios C., Rausell C., Abascal F.,
 RA Baetolli U., Fernandez J.M., Jimenez L., Postigo M., Silva P.J.,
 RA Tanames J., Viguera E., Latorre A., Valencia A., Moran F., Moya A.;
 RA "Reductive genome evolution in Buchnera aphidicola."
 RT Proc. Natl. Acad. Sci. U.S.A. 100:581-586 (2003).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.

| | | | |
|----|----------|------------------------------------|------------|
| PT | TRANSMEM | 502 | POTENTIAL. |
| PT | TRANSMEM | 539 | POTENTIAL. |
| PT | TRANSMEM | 555 | POTENTIAL. |
| SQ | SEQUENCE | 743 AA; 83644 MW; 84B9945E56B82F15 | CRC64; |

Query Match Score 33; DB 1; Length 743;
 Best Local Similarity 75.0%; Pred. No. 41;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY      3  WRADAH  10
      ||| |||
Db      294  VVEGGLHY  301

RESULT 12
GAAB_METHH
ID      GAAB_METHH      STANDARD;      PRT;      308 AA.
AC      Q26806;
DT      16-OCT-2001      (Rel. 40, Created)
DT      16-OCT-2001      (Rel. 40, Last sequence update)
DT      28-FEB-2003      (Rel. 41, Last annotation update)
DE      GMP synthase [glutamine-hydrolyzing] subunit B (EC 6.3.5.2) (GMP
DE      synthetase).

```

OC Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OC Methanobacterium thermoautotrophicum.

```

OK      DBF_180125187420;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Delta H;

```

RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Madsen J.E. 2004; PubMed 1571445;
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jivani N., Caruso A., Bueh D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shiner G., Goyal A., Pistrovski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT *Complete genome sequence of *Methanobacterium thermoautotrophicum*
RT delah: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).

CC -I- CATALYTIC ACTIVITY: ATP + xanthosine 5'-phosphate + L-glutamine +
CC H₂O → AMP + diphosphate + GMP + L-glutamate.
CC
CC -I- PATHWAY: GMP biosynthesis.
CC
CC -I- SUBUNIT: HETERODIMER COMPOSED OF A GLUTAMINE AMIDOTRANSFERASE
CC SUBUNIT (A) AND A GMP SYNTHASE SUBUNIT (B) (POTENTIAL).
CC
CC -I- SIMILARITY: BELONGS TO THE GMP SYNTHASE FAMILY.
CC
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC

```
CC ENBL; AE000850; AAB85215.1; ALT_INIT.
```

| | |
|-----------------------|---|
| DR | FLAM; PRC00356; GMP_synth_C; A. |
| DR | TIGRFA05; TIGR00884; guaA_Cterm; 1. |
| KW | Ligase; GMP biosynthesis; Purine biosynthesis; ATP-binding; |
| KM | Complete proteome. |
| FT | DOMAIN 33 184 GMP-BINDING (BY SIMILARITY). |
| FT | NP BIND 29 35 ATP (BY SIMILARITY). |
| FT | SEQUENCE 308 AA; 34403 MW; F2DCF6ED202CAEC1 CRC64; |
| SQ | |
| Query Match | 61.5%; Score 32; DB 1; Length 308; |
| Best Local Similarity | 63.8%; Pred. NO. 27; |
| Matches | 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0; |
| Qy | 1 ERVVPXGXHYS 11 |

```

Db          216 EVVPSGLHES 226
|||||
RESULT 13
ALFB SHEEP
ID ALFB SHEEP STANDARD; PRT; 363 AA.
AC P52210;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Fructose-bisphosphate aldolase B (EC 4.1.2.13) (Liver-type aldolase).
GN ALDOB.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Measonephros;
RX MEDLINE=94368863; PubMed=8086469;
RA Glanquinto L., Failhoux E.A., Bezard J., Servel N., Kirszenbaum M.,
RA Cotinot C.;
RT "Cloning and characterization of a full-length cDNA coding for ovine
RT aldolase B from fetal measonephros.";
RL Biochim. Biophys. Acta 1219:223-227(1994).
CC -1- CATALYTIC ACTIVITY: D-fructose 1,6-bisphosphate = glycerone
CC phosphate + D-glyceraldehyde 3-phosphate.
CC -1- PATHWAY: Glycolysis; sixth step.
CC -1- SUBUNIT: Homotrimer (By similarity).
CC -1- MISCELLANEOUS: In vertebrates, three forms of this ubiquitous
CC glycolytic enzyme are found, aldolase A in muscle, aldolase B in
CC liver and aldolase C in brain.
CC -1- SIMILARITY: Belongs to class I fructose-bisphosphate aldolase
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z29372; CAA82563.1; -
CC PIR; S47540; S47540.
CC HSP; P00883; IADO.
CC InterPro: IPR000741; Aldolase_I.
CC Pfam; PF00274; Glycolytic_ery; 1.
CC ProDom; PD001128; Aldolase_I; 1.
CC PROSITE; PS00158; ALDOLASE_CLASS_I; 1.
CC Lyase; Schiff base; Glycolysis; Multigene family.
CC INIT MET 0 BY SIMILARITY.
CC FT BINDING 55 55 C-1-PHOSPHATE GROUP OF THE SUBSTRATE.
CC FT BINDING 146 146 C-1-PHOSPHATE GROUP OF THE SUBSTRATE.
CC FT BINDING 229 229 SCHIFF-BASE WITH DIHYDROXYACETONE-P.
CC FT ACT_SITE 363 363 ESSENTIAL FOR ENHANCED ACTIVITY OF THE
CC ENZYME TOWARD FRUCTOSE 1,6-BISPHOSPHATE
CC AS COMPARED WITH FRUCTOSE 1-PHOSPHATE.
CC
CC SEQUENCE 363 AA; 39500 MW; FC8B45666821E28D CRC64;
Query Match: 61.5%; Score 32; DB 1; Length 363;
Best Local Similarity 62.5%; Pred. No. 32;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 2 EVVPSGXH 9
|||||
Db 189 EVIPDGSX 196
|||||
RESULT 14
SUOX_DROME

```

```

ID SUOX_DROME STANDARD; PRT; 573 AA.
AC Q9VWF4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable sulfite oxidase, mitochondrial precursor (EC 1.8.3.1).
GN CG7280.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.P., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Balles R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Beriman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahike C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foeller C., Gabrielson A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Hostin N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirekas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley; TISSUE=Embryo;
RX MEDLINE=22426066; PubMed=12537569;
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,
RA Rubin G.M., Ceiniker S.E.;
RT "A Drosophila full-length cDNA resource.";
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
CC -1- CATALYTIC ACTIVITY: Sulfite + O(2) + H(2)O = sulfate + H(2)O(2).
CC -1- COFACTOR: Molybdenum (molybdopterin) and one protoheme group (By
CC similarity).
CC -1- PATHWAY: Terminal reaction in the oxidative degradation of sulfur-
CC containing amino acids. It uses cytochrome c as an electron
CC acceptor.
CC -1- SUBCELLULAR LOCATION: Mitochondrial intermembrane space (By
CC similarity).
CC -1- SIMILARITY: WITH CYTOCHROME B5 AND NITRATE REDUCTASE.
CC -1- SIMILARITY: Contains 1 cytochrome b5 heme-binding domain.

```


Search completed: December 22, 2003, 17:42:28
Job time : 5.6 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 17:27:26 ; Search time 25.2 Seconds
(without alignments)
112.642 Million cell updates/sec

Title: US-09-909-164-42
Perfect score: 52
Sequence: 1 BEVVPXGXHYS 11

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|-------|-------|--------|----|--------|---------------------|
| 1 | 41 | 78.8 | 413 | 11 | Q8K289 | Q8K289 mus musculus |
| 2 | 36 | 69.2 | 208 | 2 | Q46486 | Q46486 corynebacte |
| 3 | 36 | 69.2 | 1057 | 16 | Q8CPJ4 | Q8CPJ4 staphylococ |
| 4 | 35 | 67.3 | 233 | 10 | Q40479 | Q40479 nicotiana t |
| 5 | 35 | 67.3 | 237 | 10 | Q9LW50 | Q9LW50 nicotiana s |
| 6 | 35 | 67.3 | 285 | 16 | Q98HU6 | Q98HU6 rhizobium l |
| 7 | 35 | 67.3 | 308 | 16 | Q9X2E2 | Q9X2E2 thermotoga |
| 8 | 35 | 67.3 | 317 | 9 | Q38317 | Q38317 lactobacill |
| 9 | 35 | 67.3 | 360 | 17 | Q27679 | Q27679 methanobact |
| 10 | 35 | 67.3 | 678 | 12 | Q9E1X6 | Q9E1X6 cercopithe |
| 11 | 35 | 67.3 | 855 | 5 | Q8IME6 | Q8IME6 drosophila |
| 12 | 35 | 67.3 | 1028 | 16 | Q9VJ11 | Q9VJ11 bruceella me |
| 13 | 35 | 67.3 | 1044 | 16 | Q8DIH0 | Q8DIH0 synecchococ |
| 14 | 34 | 65.4 | 279 | 16 | Q9RXN9 | Q9RXN9 deinococcus |
| 15 | 34 | 65.4 | 322 | 17 | Q9HLH8 | Q9HLH8 thermoplas |
| 16 | 34 | 65.4 | 495 | 5 | O16912 | O16912 caenorhabdi |

| | | | | | | |
|----|----|------|------|----|--------|--------------------|
| 17 | 34 | 65.4 | 534 | 17 | O29966 | O29966 archaeoglob |
| 18 | 34 | 65.4 | 822 | 2 | Q9ZHG7 | Q9ZHG7 streptococc |
| 19 | 34 | 65.4 | 822 | 16 | Q8E4U1 | Q8E4U1 streptococc |
| 20 | 34 | 65.4 | 822 | 16 | Q8DZ81 | Q8DZ81 streptococc |
| 21 | 34 | 65.4 | 823 | 16 | Q8N282 | Q8N282 streptococc |
| 22 | 34 | 65.4 | 823 | 16 | Q8K5Q1 | Q8K5Q1 streptococc |
| 23 | 34 | 65.4 | 825 | 16 | Q99XV4 | Q99XV4 streptococc |
| 24 | 34 | 65.4 | 962 | 5 | Q8IB11 | Q8IB11 plasmodium |
| 25 | 34 | 65.4 | 1442 | 17 | Q96VH5 | Q96VH5 aulotobus |
| 26 | 33 | 63.5 | 46 | 16 | Q8CYU7 | Q8CYU7 streptococc |
| 27 | 33 | 63.5 | 78 | 6 | Q9XST4 | Q9XST4 canis fami |
| 28 | 33 | 63.5 | 139 | 2 | Q57489 | Q57489 bacteroides |
| 29 | 33 | 63.5 | 156 | 16 | Q9PC35 | Q9PC35 xylella fas |
| 30 | 33 | 63.5 | 175 | 6 | Q8I033 | Q8I033 bos taurus |
| 31 | 33 | 63.5 | 215 | 6 | Q8I031 | Q8I031 bos taurus |
| 32 | 33 | 63.5 | 217 | 4 | Q00404 | Q00404 homo sapien |
| 33 | 33 | 63.5 | 252 | 16 | Q8YWP1 | Q8YWP1 anabaena sp |
| 34 | 33 | 63.5 | 297 | 6 | Q8HYX9 | Q8HYX9 bos taurus |
| 35 | 33 | 63.5 | 299 | 4 | Q9UEB9 | Q9UEB9 homo sapien |
| 36 | 33 | 63.5 | 364 | 5 | Q8IGN5 | Q8IGN5 drosophila |
| 37 | 33 | 63.5 | 412 | 5 | Q08523 | Q08523 ascaris suu |
| 38 | 33 | 63.5 | 480 | 5 | Q95NK1 | Q95NK1 plasmodium |
| 39 | 33 | 63.5 | 510 | 10 | Q9SA71 | Q9SA71 arabidopsis |
| 40 | 33 | 63.5 | 722 | 12 | Q9QTC7 | Q9QTC7 marek's dis |
| 41 | 33 | 63.5 | 748 | 12 | Q9YR01 | Q9YR01 ranid herpe |
| 42 | 33 | 63.5 | 802 | 16 | Q97QM8 | Q97QM8 streptococc |
| 43 | 33 | 63.5 | 816 | 2 | Q9AHT9 | Q9AHT9 streptococc |
| 44 | 33 | 63.5 | 819 | 2 | Q9ANY3 | Q9ANY3 streptococc |
| 45 | 33 | 63.5 | 819 | 10 | Q9AVK4 | Q9AVK4 pisum sativ |

ALIGNMENTS

RESULT 1

Q8K289 PRELIMINARY; PRT; 413 AA.
AC Q8K289; DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Similar to expressed sequence A1987856.
GN 1110004B19RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC032195; AAH32195.1;
DR MGD; MGI:1915724; 1110004B19RIK.
SQ SEQUENCE 413 AA; 46621 MW; A9B8A1DC70CDA0D5 CRC64;

Query Match 78.8%; Score 41; DB 11; Length 413;
Best Local Similarity 63.6%; Pred. No. 2.1;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 BEVVPXGXHYS 11
||:|:|:|:|
Db 168 BEIPPLGKHYS 178

RESULT 2

Q46486 PRELIMINARY; PRT; 208 AA.
ID Q46486; AC Q46486; DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE Hypothetical 23.0 kDa protein (Gcra).
GN GCRA.

```
OS Corynebacterium xerosis, and
OS Corynebacterium striatum.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=1725, 43770;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=C.xerosis; STRAIN=M82B;
RX MEDLINE=96117603; PubMed=8559800;
RA Tauch A., Kassing F., Kalinowski J., Puhler A.;
RT "The Corynebacterium xerosis composite transposon Tn5432 consists of
RT two identical insertion sequences, designated IS1249, flanking the
RT erythromycin resistance gene ermCX.";
RL Plasmid 34:119-131(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=C.striatum; STRAIN=M82B;
RX MEDLINE=20194806; PubMed=10732668;
RA Tauch A., Krieff S., Kalinowski J., Puhler A.;
RT "The 51,409-bp R-plasmid pR10 from the multi-resistant clinical
RT isolate Corynebacterium striatum M82B is composed of DNA segments
RT initially identified in soil bacteria and in plant, animal, and human
RT pathogens.";
RL Mol. Gen. Genet. 263:1-11(2000).
DR EMBL; U21300; AAC95478.1; -.
DR EMBL; AF024666; AG03390.1; -.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 208 AA; 23012 MW; F1504BE1ECDE85A6 CRC64;

Query Match 69.2%; Score 36; DB 2; Length 208;
Best Local Similarity 50.0%; Pred. No. 12;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 EVVPXGXHY 11
Db 130 DVIPGKHYA 139

RESULT 3
Q8CPJ4 PRELIMINARY; PRT; 1057 AA.
AC Q8CPJ4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Carbamoyl-phosphate synthase large chain.
GN SE0879.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
RA Chen Z., Wen Y.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF016746; AAO4476.1; -.
KW Complete proteome.
SQ SEQUENCE 1057 AA; 117391 MW; 8944D7D8DB1CAE59 CRC64;

Query Match 69.2%; Score 36; DB 16; Length 1057;
Best Local Similarity 63.6%; Pred. No. 71;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 EVVPXGXHY 11
Db 189 KEVWSNGLHY 199

RESULT 4
Q40479 PRELIMINARY; PRT; 233 AA.
ID Q40479
AC Q40479;
```

```
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DB REEBP-2.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamiids; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BY4; TISSUE=Leaf;
RX MEDLINE=95276459; PubMed=7756828;
RA Ohme-Takagi M., Shinshi H.;
RT "Ethylene-inducible DNA binding proteins that interact with an
RT ethylene responsive element.";
RL Plant Cell 7:173-182(1995).
DR EMBL; D38126; BAA07324.1; -.
DR HSP; O80337; ZGCC.
DR TRANSPAC; T02654; -.
DR InterPro; IPR001471; TF_ERF.
DR Pfam; PF00847; AP2-domain; 1.
DR ProDom; PD001423; TF_ERF; 1.
DR SMART; SM00380; AP2; 1.
SQ SEQUENCE 233 AA; 25563 MW; 6CD16783582C0CB5 CRC64;

Query Match 67.3%; Score 35; DB 10; Length 233;
Best Local Similarity 60.0%; Pred. No. 23;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXHY 10
Db 90 QAVVPKGRHY 99

RESULT 5
Q9LW50 PRELIMINARY; PRT; 237 AA.
AC Q9LW50;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Ethylene-responsive element binding factor.
GN NSERP2.
OS Nicotiana sylvestris (Wood tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamiids; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4096;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20399450; PubMed=10945353;
RA Kitajima S., Koyama T., Ohme-Takagi M., Shinshi H., Sato F.;
RT "Characterization of gene expression of NSERPs, transcription factors
RT of basic PR genes from Nicotiana sylvestris.";
RL Plant Cell Physiol. 41:817-824(2000).
DR EMBL; AB016264; BAA97122.1; -.
DR HSP; O80337; ZGCC.
DR InterPro; IPR001471; TF_ERF.
DR Pfam; PF00847; AP2-domain; 1.
DR ProDom; PD001423; TF_ERF; 1.
DR SMART; SM00380; AP2; 1.
SQ SEQUENCE 237 AA; 26243 MW; 01BC3EBE51E46298 CRC64;

Query Match 67.3%; Score 35; DB 10; Length 237;
Best Local Similarity 60.0%; Pred. No. 23;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXHY 10
Db 94 QAVVPKGRHY 103
```

RESULT 6

Q98HU6 PRELIMINARY; PRT; 285 AA.
 AC Q98HU6;
 DT 01-OCT-2001 (TREMBLrel. 18, Created)
 DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Cytochrome c1.
 GN MLL2705.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Phyllobacteriaceae; Mesorhizobium.
 OX NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFE303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsumoto A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 RT Mesorhizobium loti";
 RL DNA Res. 7:331-338 (2000).
 DR EMBL; AP003000; BAB49770.1; --
 DR InterPro; IPR000345; CytC_heme_bind.
 DR InterPro; IPR002326; Cyt_C1.
 DR Pfam; PF02167; Cytochrome_C1; 1.
 DR PRINTS; PR00603; CYTOCHROME_C1.
 DR PROSITE; PS00190; CYTOCHROME_C; 1.
 KW Complete proteome.
 SQ SEQUENCE 285 AA; 30961 MW; 31D9CDE2711747EE CRC64;

Query Match 67.3%; Score 35; DB 16; Length 285;
 Best Local Similarity 55.6%; Pred. No. 28;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 3 VVPXGXHY 11

Db 194 VVPEGTHYN 202

RESULT 7

Q9X2E2 PRELIMINARY; PRT; 308 AA.
 AC Q9X2E2;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE FTSH protease activity modulator HPLX.
 GN TM1822.
 OS Thermotoga maritima.
 OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
 OX NCBI_TaxID=2336;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MSB8 / DSM 3109;
 RX MEDLINE=99287316; PubMed=10360571;
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
 RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
 RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
 RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from
 RT genome sequence of Thermotoga maritima";
 RL Nature 399:323-329 (1999).
 DR EMBL; AE001819; AAD36885.1; --
 DR TIGR; TM1822; --
 DR InterPro; IPR001107; Band 7.
 DR InterPro; IPR001972; Stomatin.
 DR Pfam; PF01145; Band_7; 1.

DR PRINTS; PR00721; STOMATIN.
 DR SMART; SM00244; PHB; 1.
 KW Protease; Complete proteome.
 SQ SEQUENCE 308 AA; 34778 MW; ADE03603B5101A9D CRC64;

Query Match 67.3%; Score 35; DB 16; Length 308;
 Best Local Similarity 75.0%; Pred. No. 31;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 VVPXGXHY 10

Db 41 VVPSGIHY 48

RESULT 8

Q38317 PRELIMINARY; PRT; 317 AA.
 AC Q38317;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Lysin.
 GN Lys.
 OS Lactobacillus bacteriophage phi adh.
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
 OX NCBI_TaxID=12417;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Altermann E.;
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99384014; PubMed=10452953;
 RA Altermann E., Klein J., Henrich B.;
 RT "Primary structure and features of the genome of the Lactobacillus
 RT gasseri temperate bacteriophage phi-adh.";
 RL Gene 236:333-346 (1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95138034; PubMed=7836307;
 RA Henrich B., Binshofer B., Blaesi U.;
 RT "Primary structure and functional analysis of the lysis genes of
 RT Lactobacillus gasseri bacteriophage phi-adh.";
 RL J. Bacteriol. 177:723-732 (1995).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93231538; PubMed=8472961;
 RA Frenaux C., De Antoni G., Raya R., Klaenhammer T.;
 RT "Genetic organization and sequence of the region encoding integrative
 RT functions from Lactobacillus gasseri temperate bacteriophage phi-
 RT adh.";
 RL Gene 126:61-66 (1993).
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Engel G., Altermann E., Klein J., Henrich B.;
 RT "Structure of a genome region of the Lactobacillus gasseri temperate
 RT phage phi adh covering a repressor gene and cognate promoters.";

Query Match 67.3%; Score 35; DB 9; Length 317;
 Best Local Similarity 66.7%; Pred. No. 32;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 3 VVPXGXHY 11

Db 11

```

Db      60 VVPMGYHYA 68
RESULT 9
O27679 PRELIMINARY; PRT; 360 AA.
AC O27679;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Cell division protein.
GN MTH1642
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Delta H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RA "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
DR EMBL; AE000923; AB86115.1; --
DR InterPro; IPR005140; eRF1_1;
DR InterPro; IPR005141; eRF1_2;
DR InterPro; IPR005142; eRF1_3;
DR InterPro; IPR004405; PeLa;
DR Pfam; PF03463; eRF1_1; 1;
DR Pfam; PF03464; eRF1_2; 1;
DR Pfam; PF03465; eRF1_3; 1;
DR TIGRFAMs; TIGR00111; pelota; 1;
KW Cell division; Complete proteome.
SQ SEQUENCE 360 AA; 40814 MW; 2A000CB4B3CEFA69 CRC64;

Query Match 67.3%; Score 35; DB 17; Length 360;
Best Local Similarity 45.5%; Pred. No. 36;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy      1 BEVVPXGXHY 11
Db      98 EOLVPMGSHHT 108
RESULT 10
O9E1X6 PRELIMINARY; PRT; 678 AA.
AC O9E1X6;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Hypothetical 75.9 kDa protein.
OS Cercopithecine herpesvirus 7.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=35245;
RN [1]
RP SEQUENCE FROM N.A.
RA Gray W.L., Starnes H.B., White M.W., Ashburn C.V., Mahalingam R.;
RT "Complete sequence of the Simian Varicella Virus Genome.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF275348; AG27217.1; --
KW Hypothetical protein.
SQ SEQUENCE 678 AA; 75850 MW; A17B09E30512FE3C CRC64;

Query Match 67.3%; Score 35; DB 12; Length 678;
Best Local Similarity 50.0%; Pred. No. 72;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy      1 BEVVPXGXHY 10
Db      147 EBIIPKGRY 156
RESULT 11
O8IME6 PRELIMINARY; PRT; 855 AA.
AC O8IME6;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE CG30437-PC.
GN CG30437.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Spheroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley S.M.,
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Chert J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laso P., Lei Y., Levitsky A.C., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Furi V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galie R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,

```

RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler P.,
RA Stapleton M., Strong R., Swirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of *Drosophila melanogaster* genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Mira S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Bertram B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutnick F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.B.;
RT "Annotation of *Drosophila melanogaster* genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003786; RAN16124.1; -- 39BD5A516D6312DB CRC64;
SQ SEQUENCE 855 AA; 94532 MW; 39BD5A516D6312DB CRC64;
Query Match 67.3%; Score 35; DB 5; Length 855;
Best Local Similarity 66.7%; Pred. No. 93;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 1 BEVVPXGXH 9
Db 657 DEVVPSGDH 665
RESULT 12
ID Q8YJ11 PRELIMINARY; PRT; 1028 AA.
AC Q8YJ11;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE ATP-dependent DNA helicase.
GN BMEI0275.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=16M / ATCC 23456 / Biotype 1;
RX MEDLINE=20020109; PubMed=11756688;
RA DelVecchio V.G., Kapral V., Redkar R.J., Patra G., Mujar C., Los T.,
RA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,
RA Jablonaki L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltzman E.,
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA Haselkorn R., Kypides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448 (2002).
DR EMBL; AB009470; AAL51457.1; --
DR InterPro; IPR002086; Aldehyde_dehydr.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00271; helicase_C; 1.
DR SMART; SM00490; HELIC_C; 1.
DR SMART; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
DR Complete proteome.
SQ SEQUENCE 1028 AA; 112996 MW; A752B7042572E219 CRC64;

Query Match 67.3%; Score 35; DB 16; Length 1028;
Best Local Similarity 54.5%; Pred. No. 1.le+02;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
Qy 1 BEVVPXGXHYS 11
Db 76 EKIVPPGARYS 86
RESULT 13
ID Q8DIH0 PRELIMINARY; PRT; 1044 AA.
AC Q8DIH0;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Multidrug efflux transporter.
GN TLL1618.
OS Synecococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.
OX NCBI_TaxID=32046;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-1;
RX MEDLINE=22225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA Shikpo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
RT Thermosynechococcus elongatus BP-1.";
RL DNA Res. 9:123-130 (2002)
DR EMBL; AP005374; BAC09170.1; --
RW Complete proteome.
SQ SEQUENCE 1044 AA; 113205 MW; 00E9C13F0F636D2F CRC64;
Query Match 67.3%; Score 35; DB 16; Length 1044;
Best Local Similarity 63.6%; Pred. No. 1.2e+02;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
Qy 1 BEVVPXGXHYS 11
Db 843 BEVLPNGIGYS 853
RESULT 14
ID Q9RXN9 PRELIMINARY; PRT; 279 AA.
AC Q9RXN9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein DR0271.
GN DR0271.
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eissen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577 (1999).
DR EMBL; AE001889; AAF09867.1; --

DR TIGR: DR0271; --
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 279 AA; 31140 MW; DCEAL00E0AE8831 CRC64;

Query Match 65.4%; Score 34; DB 16; Length 279;
 Best Local Similarity 75.0%; Pred. No. 46;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 VPKGXHY 11
 |||||
 DB 100 VPLGRHY 107

RESULT 15

Q9HLH8 PRELIMINARY; PRT; 322 AA.

ID Q9HLH8
 AC Q9HLH8;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE Glucose-fructose oxidoreductase related protein.
 GN TA0250.
 OS Thermoplasma acidophilum.
 OC Archaea; Euryarchaeota; Thermoplasmatata; Thermoplasmatatales;
 OC Thermoplasmataceae; Thermoplasma.
 OX NCBI_TaxID=2303;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DSM 1728;
 RX MEDLINE=20479972; PubMed=11029001;
 RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
 RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
 RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
 acidophilum.";
 RL Nature 407:508-513 (2000).
 DR EMBL: AL445063; CAC11395.1; --
 DR InterPro: IPR000683; GFO_IDH_MocA.
 DR Pfam: PF01408; GFO_IDH_MocA_1.
 KW Complete proteome.
 SQ SEQUENCE 322 AA; 36918 MW; B8C239E71009D167 CRC64;

Query Match 65.4%; Score 34; DB 17; Length 322;
 Best Local Similarity 75.0%; Pred. No. 53;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPKXHY 10
 |||||
 DB 66 VVPDGLHY 73

Search completed: December 22, 2003, 17:51:31
 Job time : 26.2667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 22, 2003, 16:41:00 ; Search time 32.4667 Seconds
(without alignments)
53.778 Million cell updates/sec

Title: US-09-909-164-43

Perfect score: 48

Sequence: 1 BEVVPKXSYS 11

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_19Jun03:*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|----------------------------|
| 1 | 45 | 93.8 | 11 | 23 | ABB80544 Hepatitis C virus |
| 2 | 45 | 93.8 | 11 | 23 | ABB80545 Hepatitis C virus |
| 3 | 45 | 93.8 | 11 | 23 | ABB80549 Hepatitis C virus |
| 4 | 45 | 93.8 | 11 | 23 | ABB80552 Hepatitis C virus |
| 5 | 45 | 93.8 | 11 | 23 | ABB80553 Hepatitis C virus |
| 6 | 44 | 91.7 | 11 | 23 | ABB80521 Hepatitis C virus |
| 7 | 44 | 91.7 | 11 | 23 | ABB80522 Hepatitis C virus |
| 8 | 44 | 91.7 | 11 | 23 | ABB80525 Hepatitis C virus |
| 9 | 44 | 91.7 | 11 | 23 | ABB80526 Hepatitis C virus |

| | | | | | |
|----|----|------|----|----|----------------------------|
| 10 | 44 | 91.7 | 11 | 23 | ABB80530 Hepatitis C virus |
| 11 | 44 | 91.7 | 11 | 23 | ABB80535 Hepatitis C virus |
| 12 | 44 | 91.7 | 11 | 23 | ABB80536 Hepatitis C virus |
| 13 | 44 | 91.7 | 11 | 23 | ABB80539 Hepatitis C virus |
| 14 | 44 | 91.7 | 11 | 23 | ABB80540 Hepatitis C virus |
| 15 | 44 | 91.7 | 11 | 23 | ABB80559 Hepatitis C virus |
| 16 | 44 | 91.7 | 11 | 23 | ABB80563 Hepatitis C virus |
| 17 | 44 | 91.7 | 11 | 23 | ABB80564 Hepatitis C virus |
| 18 | 44 | 91.7 | 11 | 23 | ABB80565 Hepatitis C virus |
| 19 | 44 | 91.7 | 11 | 23 | ABB80566 Hepatitis C virus |
| 20 | 44 | 91.7 | 11 | 23 | ABB80567 Hepatitis C virus |
| 21 | 44 | 91.7 | 11 | 23 | ABB80568 Hepatitis C virus |
| 22 | 41 | 85.4 | 11 | 23 | ABB80547 Hepatitis C virus |
| 23 | 41 | 85.4 | 11 | 23 | ABB80548 Hepatitis C virus |
| 24 | 41 | 85.4 | 11 | 23 | ABB80551 Hepatitis C virus |
| 25 | 41 | 85.4 | 11 | 23 | ABB80556 Hepatitis C virus |
| 26 | 41 | 85.4 | 11 | 23 | ABB80557 Hepatitis C virus |
| 27 | 40 | 83.3 | 11 | 23 | ABB80524 Hepatitis C virus |
| 28 | 40 | 83.3 | 11 | 23 | ABB80528 Hepatitis C virus |
| 29 | 40 | 83.3 | 11 | 23 | ABB80529 Hepatitis C virus |
| 30 | 40 | 83.3 | 11 | 23 | ABB80533 Hepatitis C virus |
| 31 | 40 | 83.3 | 11 | 23 | ABB80534 Hepatitis C virus |
| 32 | 40 | 83.3 | 11 | 23 | ABB80538 Hepatitis C virus |
| 33 | 40 | 83.3 | 11 | 23 | ABB80542 Hepatitis C virus |
| 34 | 40 | 83.3 | 11 | 23 | ABB80543 Hepatitis C virus |
| 35 | 40 | 83.3 | 11 | 23 | ABB80546 Hepatitis C virus |
| 36 | 40 | 83.3 | 11 | 23 | ABB80550 Hepatitis C virus |
| 37 | 40 | 83.3 | 11 | 23 | ABB80554 Hepatitis C virus |
| 38 | 40 | 83.3 | 11 | 23 | ABB80555 Hepatitis C virus |
| 39 | 40 | 83.3 | 11 | 23 | ABB80561 Hepatitis C virus |
| 40 | 40 | 83.3 | 11 | 23 | ABB80562 Hepatitis C virus |
| 41 | 39 | 81.2 | 11 | 23 | ABB80527 Hepatitis C virus |
| 42 | 39 | 81.2 | 11 | 23 | ABB80527 Hepatitis C virus |
| 43 | 39 | 81.2 | 11 | 23 | ABB80531 Hepatitis C virus |
| 44 | 39 | 81.2 | 11 | 23 | ABB80532 Hepatitis C virus |
| 45 | 39 | 81.2 | 11 | 23 | ABB80537 Hepatitis C virus |

ALIGNMENTS

RESULT 1

ABB80544

ID ABB80544 standard; peptide; 11 AA.

AC ABB80544;

DT 08-OCT-2002 (first entry)

DB Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #24.

KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide; virucide.

OS Synthetic.

PH Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"

FT Modified-site 11 /note= "C-terminal amide"

FT Modified-site 11 /note= "C-terminal amide"

XX WO200208251-A2.

XX 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US23169.

XX 21-JUL-2000; 2000US-220101P.

XX

PA (CORV-) CORVAS INT INC.
 XX Lim-wilby M, Levy OE, Brunck TK;
 XX WPI; 2002-361643/39.
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C
 PT virus protease -
 XX Claim 17; Page 65; 69pp; English.
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.
 XX Sequence 11 AA;
 SQ
 Query Match 93.8%; Score 45; DB 23; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.011;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 BEVVPXGXSYS 11
 DB 1 BEVVPXGTSYS 11
 RESULT 2
 ABB80545
 ID ABB80545 standard; peptide; 11 AA.
 AC ABB80545;
 XX 08-OCT-2002 (first entry)
 XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #25.
 DE Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.
 XX Synthetic.
 OS
 FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 residue 7"
 FT Misc-difference 9 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"
 FT WO200208251-A2.
 XX 31-JAN-2002.
 XX 19-JUL-2001; 2001WO-US23169.
 XX 21-JUL-2000; 2000US-220101P.
 XX (CORV-) CORVAS INT INC.
 XX Lim-wilby M, Levy OE, Brunck TK;
 XX WPI; 2002-361643/39.
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C

PT virus protease -
 XX Claim 17; Page 65; 69pp; English.
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.
 XX Sequence 11 AA;
 SQ
 Query Match 93.8%; Score 45; DB 23; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.011;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 BEVVPXGXSYS 11
 DB 1 BEVVPXGTSYS 11
 RESULT 3
 ABB80549
 ID ABB80549 standard; peptide; 11 AA.
 AC ABB80549;
 XX 08-OCT-2002 (first entry)
 XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #29.
 DE Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.
 XX Synthetic.
 OS
 FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 residue 7"
 FT Misc-difference 9 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"
 FT WO200208251-A2.
 XX 31-JAN-2002.
 XX 19-JUL-2001; 2001WO-US23169.
 XX 21-JUL-2000; 2000US-220101P.
 XX (CORV-) CORVAS INT INC.
 XX Lim-wilby M, Levy OE, Brunck TK;
 XX WPI; 2002-361643/39.
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C
 PT virus protease -
 XX Claim 17; Page 65; 69pp; English.
 XX The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of

CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
SQ Sequence 11 AA;

Query Match 93.8%; Score 45; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.011;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 BEVVPXGXSYS 11
Db 1 BEVVPXGXSYS 11

RESULT 4
ABB80552
ID ABB80552 standard; peptide; 11 AA.
AC ABB80552;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #32.
XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Modified-site 11 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX
FN WO200208251-A2.
XX
PD 31-JAN-2002.
XX
PF 19-JUL-2001; 2001WO-US23169.
XX
PR 21-JUL-2000; 2000US-220101P.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Lin-wilby M, Levy OE, Brunck TK;
XX
DR WPI; 2002-361643/39.
XX
PT Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
PS Claim 17; Page 65; 69pp; English.

The sequence represents a peptide compound of the invention having
hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
invention are alpha-ketoamide peptide analogues. The peptides have
virucide activity, and are useful for treating and in the manufacture of
a medicament to treat disorders associated with HCV protease. A
pharmaceutical composition comprising the peptide as an active ingredient
is useful for treating disorders associated with hepatitis C virus.

Query Match 93.8%; Score 45; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.011;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 BEVVPXGXSYS 11
Db 1 BEVVPXGXSYS 11

Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 BEVVPXGXSYS 11
Db 1 BEVVPXGXSYS 11

RESULT 5
ABB80553
ID ABB80553 standard; peptide; 11 AA.
AC ABB80553;
XX
DT 08-OCT-2002 (first entry)
XX
DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #33.
XX
KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
KW virucide.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "N-terminal acetyl"
FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
FT residue 7"
FT Misc-difference 8 /note= "D-form residue"
FT Modified-site 9 /note= "D-form residue"
FT Modified-site 11 /note= "D-form residue"
FT Modified-site 11 /note= "C-terminal amide"
XX
FN WO200208251-A2.
XX
PD 31-JAN-2002.
XX
PF 19-JUL-2001; 2001WO-US23169.
XX
PR 21-JUL-2000; 2000US-220101P.
XX
PA (CORV-) CORVAS INT INC.
XX
PI Lin-wilby M, Levy OE, Brunck TK;
XX
DR WPI; 2002-361643/39.
XX
PT Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
PS Claim 17; Page 65; 69pp; English.

The sequence represents a peptide compound of the invention having
hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
invention are alpha-ketoamide peptide analogues. The peptides have
virucide activity, and are useful for treating and in the manufacture of
a medicament to treat disorders associated with HCV protease. A
pharmaceutical composition comprising the peptide as an active ingredient
is useful for treating disorders associated with hepatitis C virus.

Query Match 93.8%; Score 45; DB 23; Length 11;
Best Local Similarity 90.9%; Pred. No. 0.011;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 BEVVPXGXSYS 11
Db 1 BEVVPXGXSYS 11

RESULT 6
 ABB80521
 ID ABB80521 standard; peptide; 11 AA.
 XX
 XX AC ABB80521;
 XX
 DT 08-OCT-2002 (first entry)
 XX
 DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #1.
 XX
 KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Modified-site 11 /note= "C-terminal amide"
 FT
 FT WO200208251-A2.
 XX
 XX 31-JAN-2002.
 XX
 PF 19-JUL-2001; 2001WO-US23169.
 XX
 PR 21-JUL-2000; 2000US-220101P.
 XX
 PA (CORV-) CORVAS INT INC.
 XX
 PI Lim-wilby M, Levy OE, Brunck TK;
 XX
 DR WPI; 2002-361643/39.
 XX
 DR Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C
 PT virus protease -
 XX
 XX Claim 17; Page 64; 69pp; English.
 XX
 CC The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.
 XX
 SQ Sequence 11 AA;
 Query Match 91.7%; Score 44; DB 23; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.018;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Oy 1 BEVVPXGXSYS 11
 |||||
 Db 1 BEVVPXGXSYS 11
 RESULT 7
 ABB80522
 ID ABB80522 standard; peptide; 11 AA.
 XX
 XX AC ABB80522;
 XX
 DT 08-OCT-2002 (first entry)
 XX
 DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #2.

XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Modified-site 1 /note= "N-terminal acetyl"
 FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
 FT residue 7"
 FT Misc-difference 9 /note= "D-form residue"
 FT Modified-site 11 /note= "C-terminal amide"
 FT
 FT WO200208251-A2.
 XX
 XX 31-JAN-2002.
 XX
 PF 19-JUL-2001; 2001WO-US23169.
 XX
 PR 21-JUL-2000; 2000US-220101P.
 XX
 PA (CORV-) CORVAS INT INC.
 XX
 PI Lim-wilby M, Levy OE, Brunck TK;
 XX
 DR WPI; 2002-361643/39.
 XX
 DR Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C
 PT virus protease -
 XX
 XX Claim 17; Page 64; 69pp; English.
 XX
 CC The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.
 XX
 SQ Sequence 11 AA;
 Query Match 91.7%; Score 44; DB 23; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.018;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Oy 1 BEVVPXGXSYS 11
 |||||
 Db 1 BEVVPXGXSYS 11
 RESULT 8
 ABB80525
 ID ABB80525 standard; peptide; 11 AA.
 XX
 XX AC ABB80525;
 XX
 DT 08-OCT-2002 (first entry)
 XX
 DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #5.
 XX
 KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 KW virucide.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Modified-site 1

XX PA (CORV-) CORVAS INT INC.
XX PI Lim-wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
XX Claim 17; Page 64; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
XX Sequence 11 AA;
XX
XX Query Match 91.7%; Score 44; DB 23; Length 11;
XX Best Local Similarity 90.9%; Pred. No. 0.018; 1; Indels 0; Gaps 0;
XX Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 EEVVPXGXSYS 11
XX |||||
XX 1 EEVVPXGGSYS 11
XX
XX
XX RESULT 11
XX ABB80535
XX ID ABB80535 standard; peptide; 11 AA.
XX AC ABB80535;
XX
XX DT 08-OCT-2002 (first entry)
XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #15.
XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX OS Synthetic.
XX
XX FH Key Location/Qualifiers
XX FT Modified-site 1 /note= "N-terminal acetyl"
XX FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
XX residue 7"
XX FT Modified-site 11 /note= "C-terminal amide"
XX
XX PN WO200208251-A2.
XX
XX PD 31-JAN-2002.
XX
XX PF 19-JUL-2001; 2001WO-US23169.
XX
XX PR 21-JUL-2000; 2000US-220101P.
XX
XX PA (CORV-) CORVAS INT INC.
XX
XX PI Lim-wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX

XX PS Claim 17; Page 64; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A
CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.
XX
XX Sequence 11 AA;
XX
XX Query Match 91.7%; Score 44; DB 23; Length 11;
XX Best Local Similarity 90.9%; Pred. No. 0.018; 1; Indels 0; Gaps 0;
XX Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 EEVVPXGXSYS 11
XX |||||
XX 1 EEVVPXGGSYS 11
XX
XX
XX RESULT 12
XX ABB80536
XX ID ABB80536 standard; peptide; 11 AA.
XX AC ABB80536;
XX
XX DT 08-OCT-2002 (first entry)
XX DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #16.
XX KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
XX virucide.
XX OS Synthetic.
XX
XX FH Key Location/Qualifiers
XX FT Modified-site 1 /note= "N-terminal acetyl"
XX FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with
XX residue 7"
XX FT Misc-difference 9 /note= "D-form residue"
XX FT Modified-site 11 /note= "C-terminal amide"
XX
XX PN WO200208251-A2.
XX
XX PD 31-JAN-2002.
XX
XX PF 19-JUL-2001; 2001WO-US23169.
XX
XX PR 21-JUL-2000; 2000US-220101P.
XX
XX PA (CORV-) CORVAS INT INC.
XX
XX PI Lim-wilby M, Levy OE, Brunck TK;
XX WPI; 2002-361643/39.
XX Novel peptide compound having hepatitis C virus protease inhibitory
PT activity useful for treating disorders associated with hepatitis C
PT virus protease -
XX
XX PS Claim 17; Page 64; 69pp; English.
XX
XX The sequence represents a peptide compound of the invention having
CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
CC invention are alpha-ketoamide peptide analogues. The peptides have
CC virucide activity, and are useful for treating and in the manufacture of
CC a medicament to treat disorders associated with HCV protease. A

CC pharmaceutical composition comprising the peptide as an active ingredient
CC is useful for treating disorders associated with hepatitis C virus.

XX Sequence 11 AA;

Query Match 91.7%; Score 44; DB 23; Length 11;

Best Local Similarity 90.9%; Pred. No. 0.018; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXGXSYS 11

DB 1 BEVVPXGQSYS 11

RESULT 13

ABB80539
ID ABB80539 standard; peptide; 11 AA.

AC ABB80539;

DT 08-OCT-2002 (first entry)

DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #19.

XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;

KW virucide.

OS Synthetic.

XX Key Location/Qualifiers

FT Modified-site 1

FT Modified-site 6 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"

FT Misc-difference 8

FT Modified-site 11 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

PN WO200208251-A2.

XX 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US23169.

XX 21-JUL-2000; 2000US-220101P.

XX (CORV-) CORVAS INT INC.

PI Lim-wilby M, Levy OE, Brunck TK;

XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory activity useful for treating disorders associated with hepatitis C virus protease -

PS Claim 17; Page 65; 69pp; English.

XX The sequence represents a peptide compound of the invention having hepatitis C virus (HCV) protease inhibitory activity. The peptides of the invention are alpha-ketoamide peptide analogues. The peptides have virucide activity, and are useful for treating and in the manufacture of a medicament to treat disorders associated with HCV protease. A pharmaceutical composition comprising the peptide as an active ingredient is useful for treating disorders associated with hepatitis C virus.

XX Sequence 11 AA;

Query Match

Best Local Similarity 91.7%; Score 44; DB 23; Length 11;

Mismatches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXGXSYS 11

DB 1 BEVVPXGQSYS 11

RESULT 14

ABB80540
ID ABB80540 standard; peptide; 11 AA.

AC ABB80540;

XX 08-OCT-2002 (first entry)

XX Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #20.

XX Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;

KW virucide.

OS Synthetic.

XX Key Location/Qualifiers

FT Modified-site 1 /note= "N-terminal acetyl"

FT Modified-site 6 /note= "Norvalyl carbonyl forming keto-amide linkage with residue 7"

FT Misc-difference 8 /note= "D-form residue"

FT Misc-difference 9 /note= "D-form residue"

FT Modified-site 11 /note= "C-terminal amide"

XX WO200208251-A2.

XX 31-JAN-2002.

XX 19-JUL-2001; 2001WO-US23169.

XX 21-JUL-2000; 2000US-220101P.

XX (CORV-) CORVAS INT INC.

PI Lim-wilby M, Levy OE, Brunck TK;

XX WPI; 2002-361643/39.

XX Novel peptide compound having hepatitis C virus protease inhibitory activity useful for treating disorders associated with hepatitis C virus protease -

PS Claim 17; Page 65; 69pp; English.

XX The sequence represents a peptide compound of the invention having hepatitis C virus (HCV) protease inhibitory activity. The peptides of the invention are alpha-ketoamide peptide analogues. The peptides have virucide activity, and are useful for treating and in the manufacture of a medicament to treat disorders associated with HCV protease. A pharmaceutical composition comprising the peptide as an active ingredient is useful for treating disorders associated with hepatitis C virus.

XX Sequence 11 AA;

Query Match

Best Local Similarity 91.7%; Score 44; DB 23; Length 11;

Mismatches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXGXSYS 11

DB 1 BEVVPXGQSYS 11

RESULT 15

ABB80559
 ID ABB80559 standard; peptide; 11 AA.
 XX
 AC ABB80559;
 XX
 DT 08-OCT-2002 (first entry)
 XX
 DE Hepatitis C virus NS3/NS4a serine protease inhibitor peptide #39.
 XX
 KW Hepatitis C virus; HCV; serine protease; inhibitor; alpha-ketoamide;
 virucide.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Modified-site 1
 FT /note= "N-terminal acetyl"
 FT Modified-site 6
 FT /note= "Norvalyl carbonyl forming keto-amide linkage with
 residue 7"
 FT Misc-difference 8
 FT /note= "D-form residue"
 FT Modified-site 8
 FT /note= "Oxymethionine"
 FT Modified-site 11
 FT /note= "C-terminal amide"
 FT
 XX WO200208251-A2.
 PN
 XX
 XX 31-JAN-2002.
 PD
 XX
 XX 19-JUL-2001; 2001WO-US23169.
 PF
 XX
 XX 21-JUL-2000; 2000US-220101P.
 PR
 XX
 XX (CORV-) CORVAS INT INC.
 PA
 XX
 XX Lim-wilby M, Levy OE, Brunck TK;
 PI
 XX WPI; 2002-361643/39.
 DR
 XX
 XX Novel peptide compound having hepatitis C virus protease inhibitory
 PT activity useful for treating disorders associated with hepatitis C
 PT virus protease -
 XX
 XX Claim 17; Page 65; 69pp; English.
 PS
 XX
 CC The sequence represents a peptide compound of the invention having
 CC hepatitis C virus (HCV) protease inhibitory activity. The peptides of the
 CC invention are alpha-ketoamide peptide analogues. The peptides have
 CC virucide activity, and are useful for treating and in the manufacture of
 CC a medicament to treat disorders associated with HCV protease. A
 CC pharmaceutical composition comprising the peptide as an active ingredient
 CC is useful for treating disorders associated with hepatitis C virus.
 XX
 SQ Sequence 11 AA;
 Query Match 91.7%; Score 44; DB 23; Length 11;
 Best Local Similarity 90.9%; Pred. No. 0.018;
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 EEVVPXGXSYS 11
 |||||
 Db 1 EEVVPXGMSYS 11
 |||||

Search completed: December 22, 2003, 17:41:02
 Job time : 32.4667 secs

QY 1 BEVVPXGXSX 10
||:| | | |
Db 1 BEISPLGWSY 10

RESULT 2

US-08-871-355A-236
; Sequence 236, Application US/08871355A
; Patent No. 6015669
; GENERAL INFORMATION:
; APPLICANT: David William Holden
; TITLE OF INVENTION: Identification of Genes
; NUMBER OF SEQUENCES: 501
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/871,355A
; FILING DATE: 09-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB95/02875
; FILING DATE: 11-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: RPMS 101 CON
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 236:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 45 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
US-08-871-355A-236

Query Match 66.7%; Score 32; DB 3; Length 45;
Best Local Similarity 60.0%; Pred. No. 8.3;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXSX 10
||:| | | |
Db 1 BEISPLGWSY 10

RESULT 3

US-09-201-945-236
; Sequence 236, Application US/09201945
; Patent No. 6342215
; GENERAL INFORMATION:
; APPLICANT: David William Holden
; TITLE OF INVENTION: Identification of Genes
; NUMBER OF SEQUENCES: 501
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; CITY: Atlanta

STATE: Georgia
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/201,945
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/637,759
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: RPMS 101
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404) 873-8794
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 236:
SEQUENCE CHARACTERISTICS:
LENGTH: 45 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-09-201-945-236

Query Match 66.7%; Score 32; DB 4; Length 45;
Best Local Similarity 60.0%; Pred. No. 8.3;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 BEVVPXGXSX 10
||:| | | |
Db 1 BEISPLGWSY 10

RESULT 4

US-09-228-986-73
; Sequence 73, Application US/09228986
; Patent No. 6359198
; GENERAL INFORMATION:
; APPLICANT: Strabala, Timothy
; APPLICANT: Nieuwenhuizen, Niels
; TITLE OF INVENTION: Compositions Isolated from Plant Cells
; TITLE OF INVENTION: and Their Use in the Modification of Plant Cell Signalling
; FILE REFERENCE: 11000/1020
; CURRENT APPLICATION NUMBER: US/09/228,986
; CURRENT FILING DATE: 1999-01-12
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: RastSEQ for Windows Version 3.0
; SEQ ID NO 73
; LENGTH: 947
; TYPE: PRT
; ORGANISM: Pinus radiata
US-09-228-986-73

Query Match 66.7%; Score 32; DB 4; Length 947;
Best Local Similarity 66.7%; Pred. No. 2.4e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXSYS 11
||:| | | |
Db 686 VMPGSGISYS 694

RESULT 5
US-08-844-086-4

Sequence 4, Application US/08844086
Patent No. 5866390

GENERAL INFORMATION:
APPLICANT: Lawlor, Elizabeth
TITLE OF INVENTION: No. 5866390el Compounds
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/844,086
FILING DATE: 18-APR-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9607993.4
FILING DATE: 18-APR-1996

ATTORNEY/AGENT INFORMATION:
NAME: Gimmi, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P31457-4

TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:
LENGTH: 159 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-844-086-4

Query Match 64.6%; Score 31; DB 2; Length 159;
Best Local Similarity 66.7%; Pred. No. 53;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 REVVPXGXS 9
DB 123 EEVLDPDGT 131

RESULT 6

US-09-018-211-4

Sequence 4, Application US/09018211
Patent No. 6048716

GENERAL INFORMATION:
APPLICANT: Lawlor, Elizabeth
TITLE OF INVENTION: No. 6048716el Compounds
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/018,211

FILING DATE:

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/844,086
FILING DATE: 18-APR-1997
APPLICATION NUMBER: 9607993.4
FILING DATE: 18-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmi, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P31457-4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 159 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-018-211-4

Query Match 64.6%; Score 31; DB 3; Length 159;
Best Local Similarity 66.7%; Pred. No. 53;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 REVVPXGXS 9
DB 123 EEVLDPDGT 131

RESULT 7

PCT-US91-02714-26

Sequence 26, Application PC/TUS9102714
GENERAL INFORMATION:

APPLICANT: Wigler, Michael H.
APPLICANT: Colicelli, John J.
TITLE OF INVENTION: Cloning by Complementation and Related
NUMBER OF SEQUENCES: 55
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
ADDRESS: Bicknell
STREET: Two First National Plaza, 20 South Clark
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60603

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US91/02714
FILING DATE: 19910419
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/511,715
FILING DATE: 20-APR-1990

ATTORNEY/AGENT INFORMATION:
NAME: Borun, Michael F.
REGISTRATION NUMBER: 25447
REFERENCE/DOCKET NUMBER: 27805/30197
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 346-5750
TELEFAX: (312) 984-9740
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:

/ LENGTH: 513 amino acids
/ TYPE: AMINO ACID
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
PCT-US91-02714-26

Query Match 64.6%; Score 31; DB 5; Length 513;
Best Local Similarity 75.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXS 10
Db 201 VVPAGGS 208

RESULT 8

US-08-459-146-2
; Sequence 2, Application US/08459146
; Patent No. 5866405
; GENERAL INFORMATION:
; APPLICANT: Choi, Gil Ho
; APPLICANT: Nuss, Donald Lee
; TITLE OF INVENTION: Genetically Engineered Transmissible
; TITLE OF INVENTION: Hypovirulence
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: George M. Gould, Esq., Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07110

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,146
FILING DATE: 02-JUN-1995
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/832,117
FILING DATE: 06-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: Roseman, Catherine R
REGISTRATION NUMBER: 34,240
TELEPHONE: (201) 235-6208
TELEFAX: (201) 235-3500
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 622 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO

ORIGINAL SOURCE:
ORGANISM: Endothia parasitica (Cryptonectria
ORGANISM: parasitica)
STRAIN: EP713
US-08-459-146-2

Query Match 64.6%; Score 31; DB 2; Length 622;
Best Local Similarity 85.7%; Pred. No. 2.4e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXG 7
Db 31 BEVVPAG 37

RESULT 9

US-08-459-065-2
; Sequence 2, Application US/08459065
; Patent No. 5882642
; GENERAL INFORMATION:
; APPLICANT: Choi, Gil Ho
; APPLICANT: Nuss, Donald Lee
; TITLE OF INVENTION: Genetically Engineered Transmissible
; TITLE OF INVENTION: Hypovirulence
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: George M. Gould, Esq., Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07110

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,065
FILING DATE: 02-JUN-1995
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/832,117
FILING DATE: 06-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: Roseman, Catherine R
REGISTRATION NUMBER: 34,240
REFERENCE/DOCKET NUMBER: 8589

TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 235-6208
TELEFAX: (201) 235-3500
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 622 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Endothia parasitica (Cryptonectria
ORGANISM: parasitica)
STRAIN: EP713
US-08-459-065-2

Query Match 64.6%; Score 31; DB 2; Length 622;
Best Local Similarity 85.7%; Pred. No. 2.4e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 BEVVPXG 7
Db 31 BEVVPAG 37

RESULT 10

US-07-688-352C-28
; Sequence 28, Application US/07688352C
; Patent No. 5527896
; GENERAL INFORMATION:
; APPLICANT: Wigler, Michael H.
; APPLICANT: Colicelli, John J.
; TITLE OF INVENTION: Cloning by Complementation and Related
; TITLE OF INVENTION: Processes
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Bicknell
; STREET: Two First National Plaza, 20 South Clark
; STREET: Street

```

; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/688,352C
; FILING DATE: 19910419
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/511,715
; FILING DATE: 20-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Borun, Michael F.
; REGISTRATION NUMBER: 25447
; REFERENCE/DOCKET NUMBER: 27805/30197
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 346-5750
; TELEFAX: (312) 984-9740
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 688 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-688-352C-28

Query Match 64.6%; Score 31; DB 1; Length 688;
Best Local Similarity 75.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXSX 10
   ||| |||
Db 201 VVPAGGSY 208

RESULT 11
US-08-474-379C-28
; Sequence 28, Application US/08474379C
; Patent No. 597305
; GENERAL INFORMATION:
; APPLICANT: Wigler, Michael H.
; APPLICANT: Colicelli, John J.
; TITLE OF INVENTION: CLONING BY COMPLEMENTATION AND RELATED
; TITLE OF INVENTION: PROCESSES
; NUMBER OF SEQUENCES: 88
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive/6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,379C
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/511,715
; FILING DATE: 20-APR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/206,188
; FILING DATE: 01-MAR-1994

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/688,352
; FILING DATE: 19-APR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Clough, David W.
; REGISTRATION NUMBER: 36,107
; REFERENCE/DOCKET NUMBER: 27866/32771
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 688 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-474-379C-28

Query Match 64.6%; Score 31; DB 2; Length 688;
Best Local Similarity 75.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 VVPXGXSX 10
   ||| |||
Db 201 VVPAGGSY 208

RESULT 12
US-09-146-249A-28
; Sequence 28, Application US/09146249A
; Patent No. 6069240
; GENERAL INFORMATION:
; APPLICANT: Wigler, Michael H.
; APPLICANT: Colicelli, John J.
; TITLE OF INVENTION: Cloning by Complementation and Related
; TITLE OF INVENTION: Processes
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/146,249A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/511,715
; FILING DATE: 20-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Clough, David W.
; REGISTRATION NUMBER: 36,107
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 688 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-146-249A-28

Query Match 64.6%; Score 31; DB 3; Length 688;
Best Local Similarity 75.0%; Pred. No. 2.7e+02;

```